Tin học trong nghiên cứu khoa học

TS.GVC. Trần Thúy Hiền
Bộ môn Toán-Tin-Sinh, Khoa Cơ bản,
ĐH Y Dược Huế
(Tầng 4, Nhà B, Số 06 Ngô Quyền, Tp Huế)

Nội dung học phần

1. NCKH – Thiết kế nghiên cứu y học
2. Tổng quan về phần mềm SPSS
3. Các thao tác cơ bản trên SPSS
4. SPSS trong Thống kê mô tả
5. SPSS trong Kiểm định giả thuyết
6. SPSS trong Tương quan và Hồi qui

Chương 1. NCKH – Thiết kế NC

1. Khái niệm NCKH

- NCKH là khảo cứu một vấn đề mới mẻ nào đó được tiến hành theo tất cả các quy tắc chứng minh.
- NCKH là hoạt động tìm kiếm, xem xét, điều tra hoặc thử nghiệm để phát hiện tìm ra những cái mới về bản chất sự vật, về thế giới tự nhiên - xã hội, và để sáng tạo ra phương pháp, phương tiện kỹ thuật mới cao hơn, giá trị hơn.

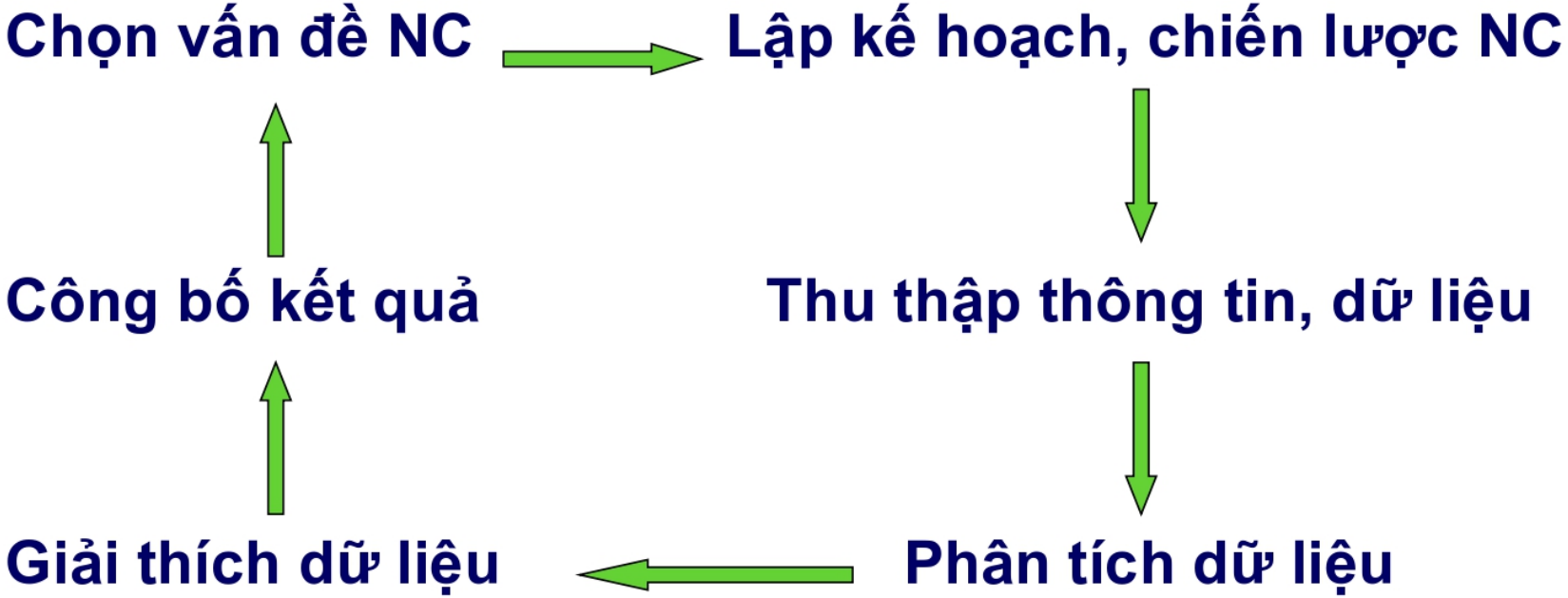
Chương 1. NCKH – Thiết kế NC

- NCKH đặc trưng bởi 2 dấu hiệu cơ bản:
  - Tính mới mẻ của vấn đề nghiên cứu
  - Tính chất có chứng minh, có lí lẽ vững chắc của chính những kết luận, luận điểm mà người nghiên cứu nêu lên.

2. Các loại hình nghiên cứu

- **NC cơ bản**: trong phòng thí nghiệm, ở tầm phân tử, tế bào và có thể sử dụng sinh vật thí nghiệm, các mô và các cơ quan của người, nhưng chưa phải trên cơ thể người.
- **NC ứng dụng**: vận dụng các quy luật được phát hiện qua NC cơ bản để tìm các giải pháp, nguyên lí của giải pháp. NC trên quy mô nhỏ.
- **NC triển khai**: áp dụng các kết quả NC ứng dụng vào thực tế phục vụ trực tiếp cho phòng bệnh, khám chữa bệnh và phục hồi chức năng. Áp dụng trên quy mô rộng rãi.

3. Quy trình một NCKH



3. Quy trình một NCKH

- Bước 1: Chọn vấn đề NC (ý tưởng NC, Tổng quan tài liệu, mục tiêu, đối tượng NC, Giả thuyết khoa học)
- **Bước 2:** Thiết kế NC (chọn mẫu NC, loại dữ liệu phải thu thập và nguồn cung cấp dữ liệu, phương pháp thu thập dữ liệu; thiết kế bảng câu hỏi; phương pháp xử lý dữ liệu)
- Bước 3: Thu thập dữ liệu
- **Bước 4:** Xử lý, phân tích dữ liệu
- **Bước 5:** Diễn giải, giải thích dữ liệu
- Bước 6: So sánh các NC trước
- Bước 7: Kết luận, trình bày, báo cáo kết quả.

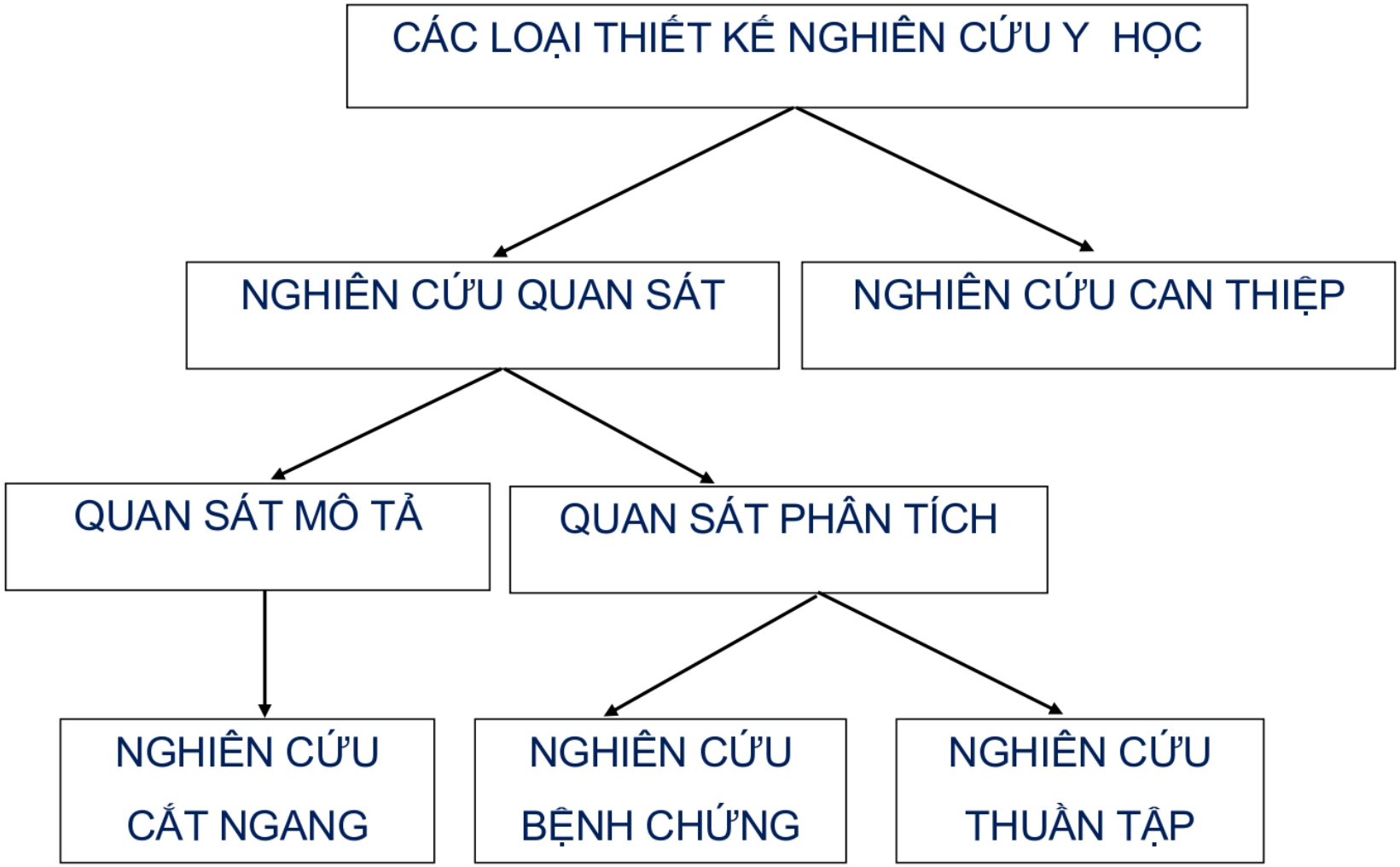
4. Thiết kế nghiên cứu Y học

Phân loại thiết kế nghiên cứu y học phổ biến nhất là cách phân loại theo cách thiết kế nghiên cứu dịch tễ học.

Cách phân loại thiết kế này thường ứng dụng cách tiếp cận trong điều tra dịch tễ học, tập trung vào đo lường độ lớn của vấn đề sức khỏe, kiểm tra mối quan hệ nhân quả của các biến số, từ đó nhằm phát hiện ra các yếu tố căn nguyên, phơi nhiễm, các yếu tố liên quan đến vấn đề sức khỏe hoặc hiệu quả của các giải pháp can thiệp.

Trần Thúy Hiền

Chương 1. NCKH – Thiết kế NC



Trần Thúy Hiền

4. Thiết kế nghiên cứu Y học

- **NC quan sát**: không tác động vào hiện tượng quan tâm, chỉ quan sát, không can thiệp.
- **NC can thiệp**: người NC can thiệp vào hoặc tạo ra một yếu tố được coi là nguyên nhân rồi theo dõi, ghi nhận kết quả của can thiệp (phân tích mối quan hệ nhân - quả, kiểm định giả thuyết nhân quả).

Chương 1. NCKH – Thiết kế NC

			Mục đích	Chọn mẫu
NC Quan sát	QS mô tả	NC cắt ngang	Hình thành giả thuyết nhân quả	Chọn 1 nhóm đối tượng
	QS phân tích	NC bệnh chứng	Kết luận về giả thuyết nhân quả	Chọn 2 nhóm đối tượng để so sánh
		NC thuần tập		
NC Can thiệp			Kiểm định về giả thuyết nhân quả	

Chương 1. NCKH – Thiết kế NC

- Quan sát mô tả: mô tả bệnh và một (hay nhiều) yếu tố nguy cơ của bệnh để tìm ra mối liên quan có thể là kết hợp nhân quả tại một thời điểm.

Mục đích: hình thành Giả thuyết nhân - quả (không chứng minh)

Ví dụ: Nghiên cứu mô tả (tương quan) cho thấy người dân tỉnh Henan Trung Quốc (với dân số gần 100 triệu người) sử dụng muối ăn nhiều và nhận thấy số bệnh nhân bị ung thư thực quản cao.

- Quan sát phân tích: quan tâm cả quá trình diễn biến mối liên hệ nhân quả, cả quan sát và phân tích, quan sát để tìm bệnh hoặc yếu tố phơi nhiễm.

Mục đích: kết luận về Giả thuyết nhân - quả.

Chương 1. NCKH – Thiết kế NC

- ***Quan sát mô tả***:
  - **NC cắt ngang (Cross sectional study):** áp dụng để mô tả hiện tượng sức khỏe và các yếu tố được cho là có liên quan đến hiện tượng sức khỏe đó của quần thể tại một thời điểm nhất định, mỗi đối tượng chỉ thu thập thông tin 1 lần.

    Sản phẩm là tỉ lệ mắc bệnh và các giả thuyết nhân quả.

    Thường sử dụng trong lâm sàng vì dễ thực hiện, góp phần nâng cao chất lượng chẩn đoán, điều trị tại bệnh viện hoặc giúp đưa ra các giả thuyết cho các nghiên cứu thực nghiệm tiếp theo.

    Cách chọn mẫu nghiên cứu: Chọn một nhóm đối tượng từ quần thể quan tâm.

- ***Quan sát phân tích***:
  - **NC bệnh chứng (Cases-control study) và NC thuần tập (Prospective Cohort study)**

    Cách chọn mẫu NC: Chọn 2 nhóm đối tượng để so sánh nguy cơ mắc bệnh.

Chương 1. NCKH – Thiết kế NC

1

Chủ động chọn 2 nhóm đối tượng mắc bệnh và không mắc bệnh trước. Sau đó hồi cứu (khai thác trong quá khứ) về việc phơi nhiễm hoặc không phơi nhiễm với yếu tố bị nghi ngờ là nguyên nhân của bệnh.

NC bệnh chứng

2

Chủ động chọn 2 nhóm đối tượng có hoặc không phơi nhiễm với yếu tố nguy cơ của bệnh. Sau đó theo dõi trong tương lai để ghi nhận sự xuất hiện bệnh.

NC thuần tập

Chương 1. NCKH – Thiết kế NC

Xác định các phương pháp nghiên cứu đối với các Đề tài NCKH sau:

1. Đề tài: Đánh giá tỷ lệ và phân loại rối loạn Lipid máu trên bệnh nhân hội chứng thận hư không đơn thuần (có tăng HA và/hoặc đái máu và/hoặc suy thận).

Đối tượng NC: 47 bệnh nhân chẩn đoán hội chứng thận hư không đơn thuần, nguyên phát tuổi từ 16 trở lên, điều trị từ 3/2014 đến 3/2015 tại khoa nội cơ xương khớp BV TW Huế.

Kết quả: 98% bệnh nhân có rối loạn ít nhất một thành phần Lipid máu, 75% bệnh nhân rối loạn cả 3 thành phần của bilan Lipid (tăng Cholesterol toàn phần, triglyceride và LDL-Cholesterol)…

Trần Thúy Hiền

Chương 1. NCKH – Thiết kế NC

2. Đề tài: Nghiên cứu nồng độ Leptin và Adiponectin huyết tương ở người thừa cân, béo phì.

Đối tượng NC: Nhóm bệnh 70 người thừa cân béo phì, nhóm chứng 67 người bình thường, xác định theo tiêu chuẩn WHO

Kết quả: Có sự tăng dần về nồng độ Leptin huyết tương và giảm dần về nồng độ Adiponectin huyết tương từ nhóm chứng đến nhóm thừa cân và nhóm béo phì….

Chương 1. NCKH – Thiết kế NC

3. Nghiên cứu thực hiện năm 1951, tại Anh

Mục đích NC: Xác định mối liên hệ nhân-quả giữa hút thuốc lá và ung thư phổi.

Đối tượng NC: 2 nhóm đối tượng (2 đoàn hệ): hút thuốc lá và không hút thuốc lá. Sau 40 năm (1951 – 1991), xác định bệnh ung thư phổi trên các đối tượng.

Kết quả: tỷ lệ bị ung thư phổi trong nhóm có hút thuốc là cao hơn so với nhóm không hút thuốc. Có sự liên hệ giữa "liều lượng – đáp ứng" nghĩa là hút càng nhiều, khả năng mắc bệnh ung thư phổi càng cao…

Hút thuốc lá là một nguyên nhân của ung thư phổi.

Thank You !
Email: tthien@huemed-univ.edu.vn

Chương 2. Tổng quan về phần mềm SPSS

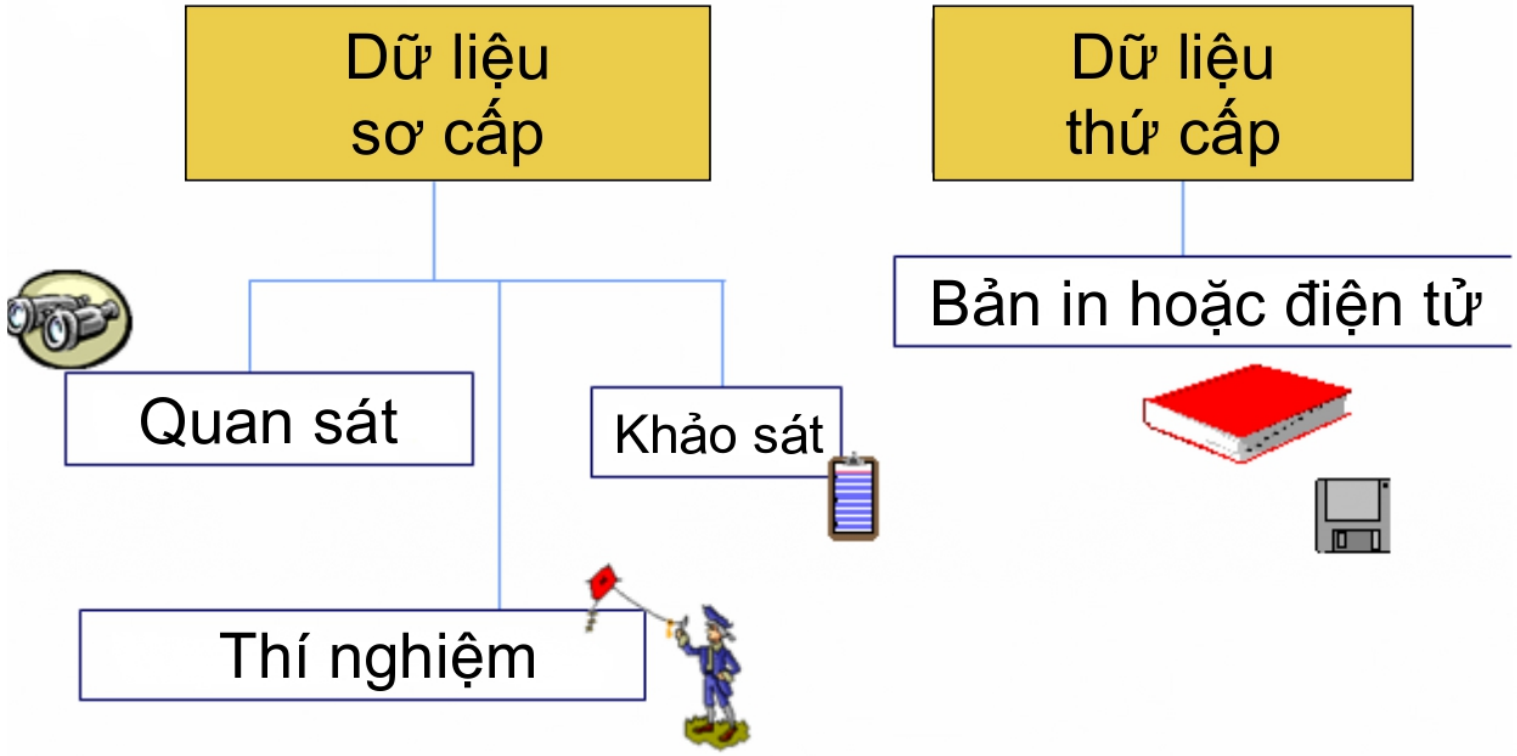
2.1. Quá trình thu thập và xử lý dữ liệu
2.2. Giới thiệu về phần mềm SPSS
2.3. Khởi tạo biến trên SPSS
2.4. Nhập dữ liệu và làm sạch dữ liệu

Trần Thúy Hiền

Chương 2. Tổng quan về phần mềm SPSS

2.1. Quá trình thu thập và xử lý dữ liệu

(1) Thu thập dữ liệu

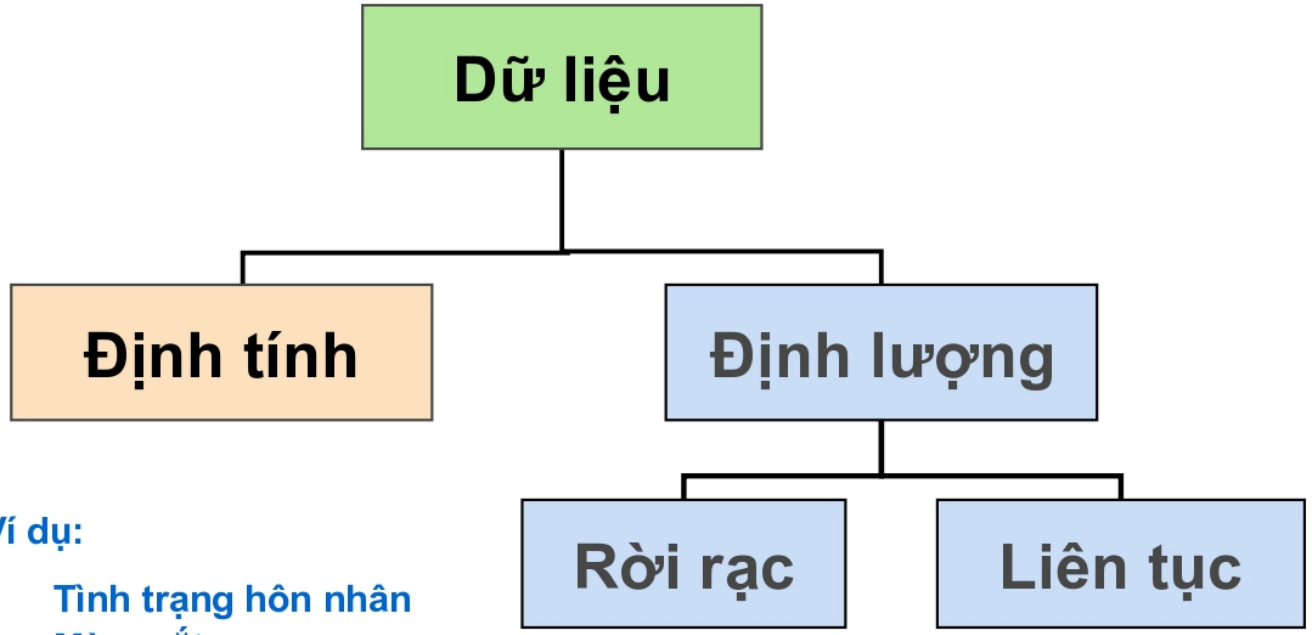


Chương 2. Tổng quan về phần mềm SPSS

- Nguồn dữ liệu sơ cấp: Người tập hợp dữ liệu cũng là người sử dụng dữ liệu để phân tích:
  - Dữ liệu từ các khảo sát
  - Dữ liệu tập hợp từ thí nghiệm
  - Dữ liệu quan sát được
- Nguồn dữ liệu thứ cấp: Người thực hiện việc phân tích dữ liệu không phải là người tập hợp dữ liệu:
  - Phân tích nguồn dữ liệu quốc gia
  - Khảo sát dữ liệu từ các ấn bản báo giấy hay từ các dữ liệu được xuất bản trên internet.

Chương 2. Tổng quan về phần mềm SPSS

(2) Các kiểu dữ liệu



Ví dụ:

- Tình trạng hôn nhân
- Màu mắt

Ví dụ:

- Số con
- Số bệnh nhân/ 1 ngày

Ví dụ:

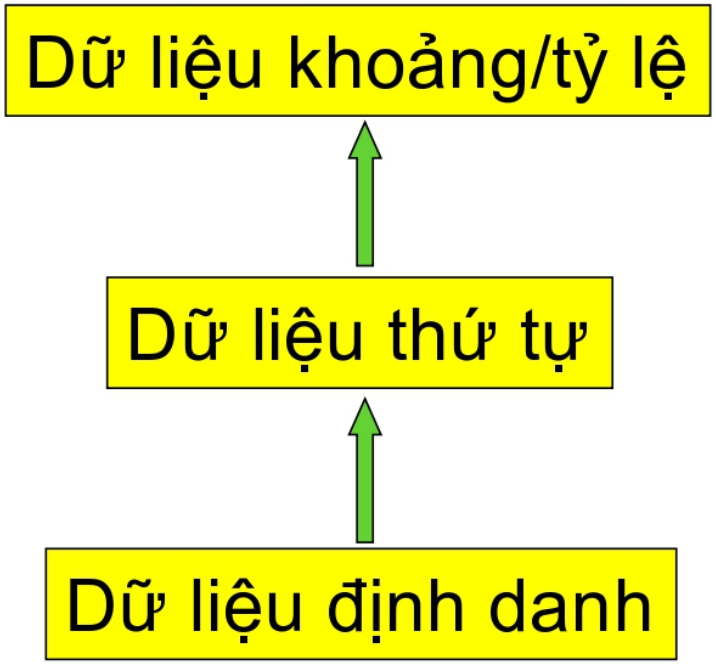
- Cân nặng
- Chiều cao

Chương 2. Tổng quan về phần mềm SPSS

- Dữ liệu định tính: phản ánh tính chất, sự hơn kém của đối tượng của các đối tượng nghiên cứu, là các dữ liệu ban đầu không thể hiện dưới dạng số;
- Dữ liệu định lượng: phản ánh mức độ, hay mức độ hơn kém của các đối tượng nghiên cứu, là các dữ liệu có thể cân đong, đo, đếm … được.

Chương 2. Tổng quan về phần mềm SPSS

(3) Các thang đo dữ liệu



Chương 2. Tổng quan về phần mềm SPSS

- **Thang đo định danh** (nominal)
  - Phân dữ liệu thành các lớp riêng biệt và không có thứ tự;
  - Dữ liệu ở mức thang đo định danh chỉ có thể là dữ liệu định tính;
  - Các con số chỉ dùng để phân loại đối tượng, không mang ý nghĩa nào khác;
  - Ví dụ: họ tên, giới tính, nghề nghiệp, nơi ở…

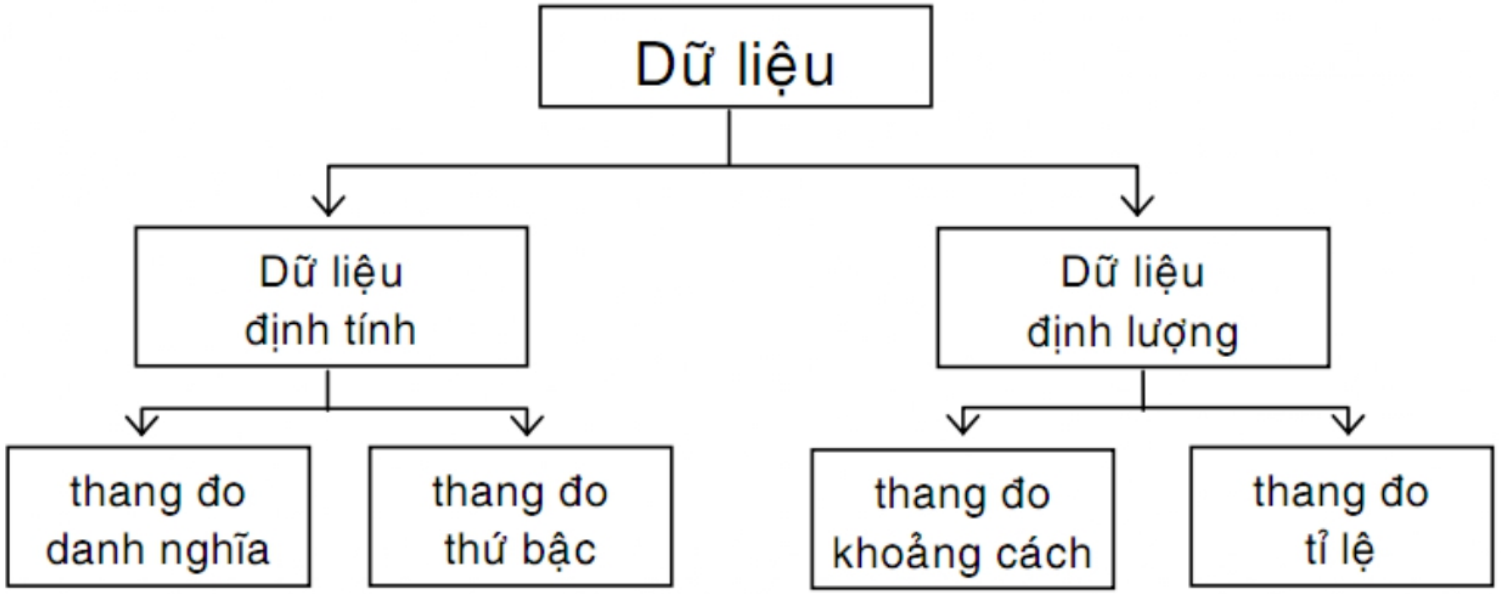
Chương 2. Tổng quan về phần mềm SPSS

- **Thang đo thứ tự** (ordinal)
  - Phân dữ liệu thành các lớp riêng biệt và có thứ tự;
  - Dữ liệu ở mức thang đo thứ tự được sắp xếp thứ bậc nhưng khoảng cách giữa các mục dữ liệu là không có ý nghĩa.
  - Ví dụ: mức độ suy dinh dưỡng của trẻ là 1, 2, 3; mức độ bỏng I, II, III

Chương 2. Tổng quan về phần mềm SPSS

- **Thang đo khoảng** (interval)
  - Là một dạng đặc biệt của thang đo thứ tự và khoảng cách giữa các giá trị là một con số có nghĩa;
  - Thang đo này không có một điểm 0 thật sự (chỉ do quy ước);
  - Ví dụ: Nhiệt độ
- **Thang đo tỷ lệ** (ratio)
  - giống như ở mức thang đo khoảng nhưng điểm 0 là có một trị số thật;
  - Cho phép lấy tỷ lệ so sánh giữa 2 giá trị.
  - Ví dụ: Chiều cao, cân nặng, tuổi, thu nhập…

Chương 2. Tổng quan về phần mềm SPSS



Trần Thúy Hiền

Chương 2. Tổng quan về phần mềm SPSS

(4) Quy trình xử lý dữ liệu

Thông tin thu thập thông qua Bảng câu hỏi

- Bước 1: Kiểm tra, hiệu chỉnh bảng câu hỏi
- Bước 2: Mã hóa dữ liệu trên bảng câu hỏi
- Bước 3: Nhập dữ liệu vào máy tính
- Bước 4: Phát hiện lỗi và làm sạch dữ liệu trên máy tính

Chương 2. Tổng quan về phần mềm SPSS

(5) Mã hóa dữ liệu

- Dữ liệu thu thập được mã hóa dưới dạng một bảng 2 chiều (dòng x cột);
  - Các cột: thể hiện các biến đang được đo lường, khảo sát; mỗi cột (biến) chứa các câu trả lời của một câu hỏi nào đó
  - Các dòng: tương ứng với kết quả trả lời của từng bản câu hỏi; mỗi dòng là kết quả trả lời của một bản câu hỏi.

Chương 2. Tổng quan về phần mềm SPSS

Các biến

Những người trả lời

	Giới tính	Tuổi	Nghề nghiệp
1	Nữ	21	Sinh viên
2	Nữ	32	Nhân viên văn phòng
3	Nam	53	Về hưu
...	...	...	...
n	Nam	42	Nghề khác

Chương 2. Tổng quan về phần mềm SPSS

- Với các dữ liệu định tính không phải ký số như "nữ" hay "nhân viên văn phòng" cần phải được tạo một con số mã hóa.
  - Ví dụ: Nam nhập "1", nữ nhập "2"
- Các thông tin thu thập dưới dạng dữ liệu định lượng đã ở dưới dạng số nên không cần mã hóa.

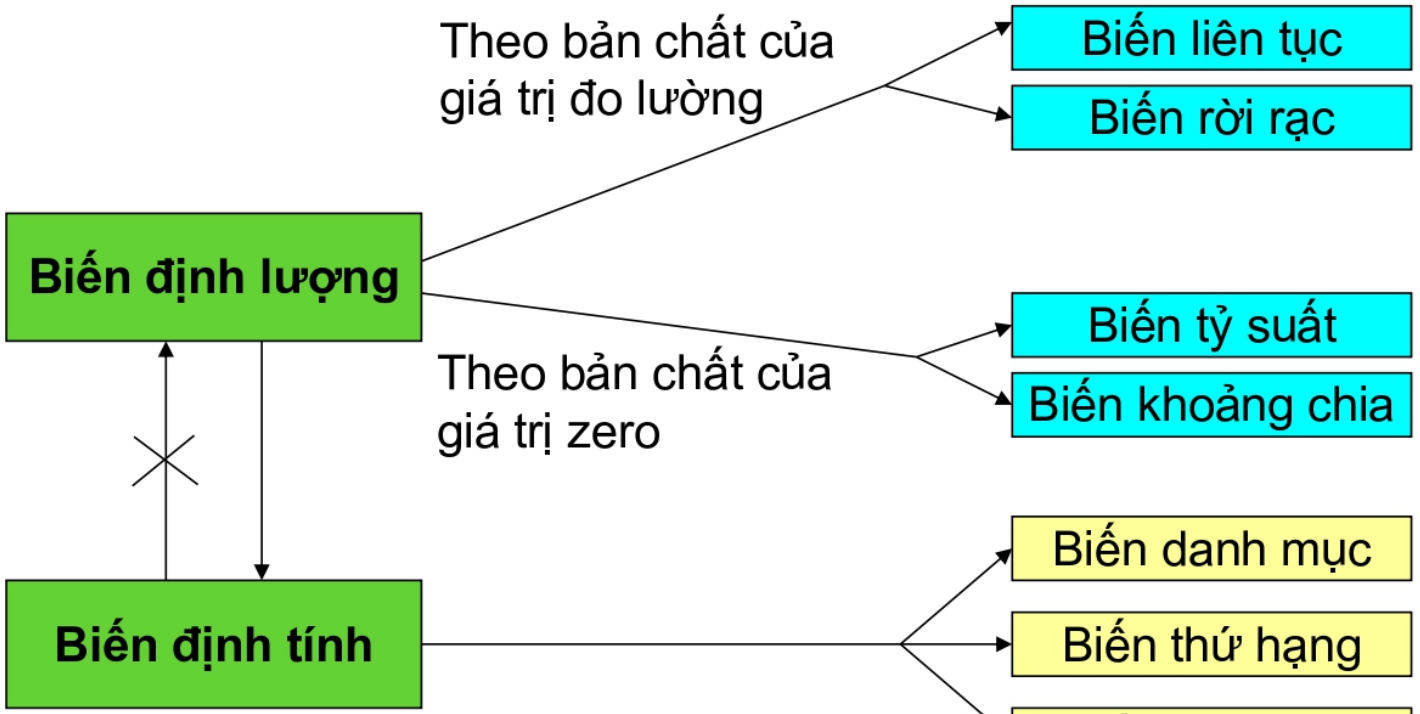
Chương 2. Tổng quan về phần mềm SPSS

(6) Biến số và phân loại biến số

- Biến số là một đặc tính của người, vật, sự việc, hiện tượng mà có thể mang các giá trị khác nhau;
- Biến số được người nghiên cứu lựa chọn để quan sát, đo lường trong quá trình nghiên cứu thì được gọi là các biến số nghiên cứu.
- Phân loại theo bản chất của biến số (theo kiểu dữ liệu)
- Phân loại theo mối tương quan giữa các biến số

Chương 2. Tổng quan về phần mềm SPSS

Phân loại theo bản chất của biến số



Chương 2. Tổng quan về phần mềm SPSS

Biến định tính	Biến định lượng
- Thể hiện tính chất - Các giá trị quan sát ở thang đo nominal, ordinal - Giá trị biến biểu thị bằng chữ hoặc ký hiệu được xếp vào các nhóm khác nhau.	-Thể hiện độ lớn - Các giá trị quan sát ở thang đo interval, ratio - Giá trị biến biểu thị bằng các con số (bản chất các con số là rời rạc hoặc liên tục).

Chương 2. Tổng quan về phần mềm SPSS

Phân loại theo mối tương quan giữa các biến số

- Biến độc lập: là biến được sử dụng để mô tả hoặc đo lường các yếu tố mà người nghiên cứu cho rằng nó có thể là nguyên nhân hoặc là yếu tố ảnh hưởng đến vấn đề đang được nghiên cứu;

  Ví dụ: Nhiễm HPV là yếu tố nguy cơ (*biến độc lập*) của bệnh ung thu cổ tủ cung chứ bệnh ung thu cổ tử cung không gây ra nhiễm HPV.

- Biến phụ thuộc: là biến được sử dụng để mô tả hoặc đo lường vấn đề cần nghiên cứu, nó có thể là hậu quả trong mối liên quan với nhiều yếu tố khác;

  Ví dụ: bướu cổ đơn thuần là một *biến phụ thuộc* vào tình trạng thiếu iod trong chế độ ăn, uống.

Chương 2. Tổng quan về phần mềm SPSS

2.2. Giới thiệu về phần mềm SPSS

SPSS: Statistical Package for Social Sciences

Cài đặt SPSS: Version 20, 23

Khởi động SPSS

Cài đặt gõ Tiếng Việt trong SPSS

Trần Thúy Hiền

CÀI ĐẶT TIẾNG VIỆT CHO SPSS

1. Khởi động SPSS, mở 1 file trống
2. Trên thanh Menu: Chọn **Edit / Options**
3. Trong **General** / Character Encoding for Data and Syntax: Chọn **Unicode**

   Chọn Apply / OK

Thoát SPSS và khởi động lại

4. Trong **Viewer:** Chọn Times New Roman

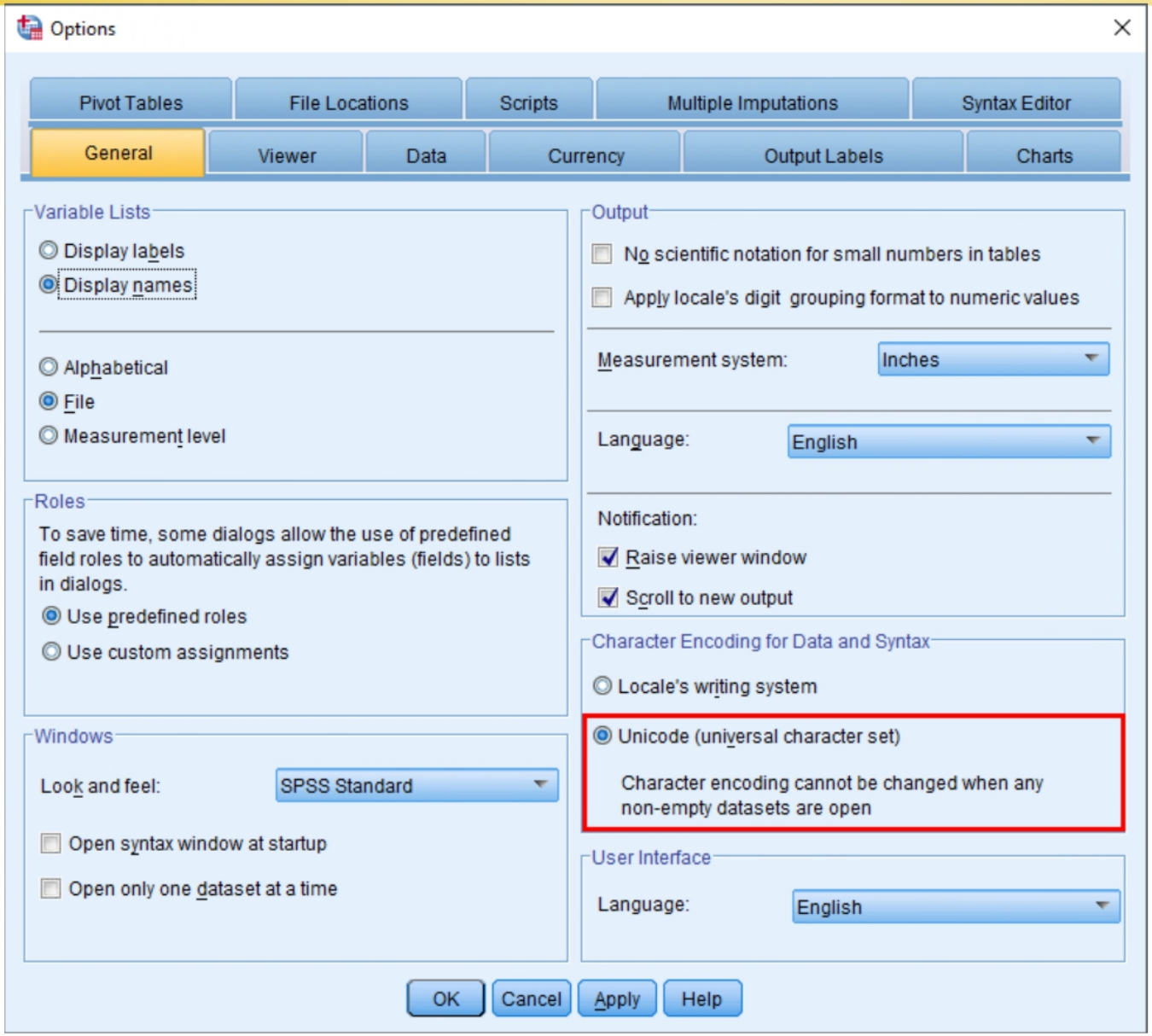
   Chọn Apply / OK

5. Trên thanh Menu: Chọn View / fonts

   Chọn Times New roman / OK

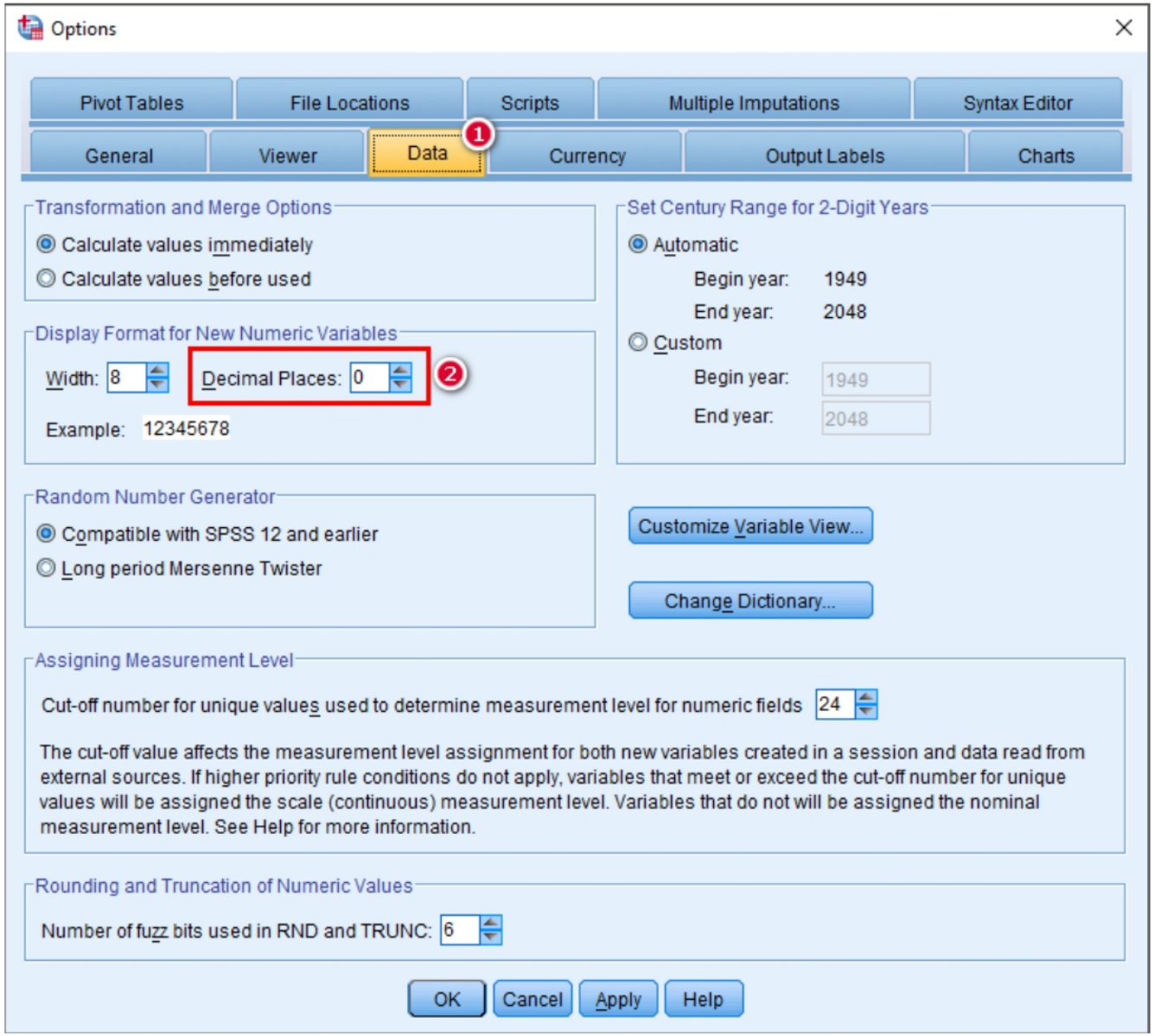
Trần Thúy Hiền

CÀI ĐẶT TIẾNG VIỆT CHO SPSS



Trần Thúy Hiền

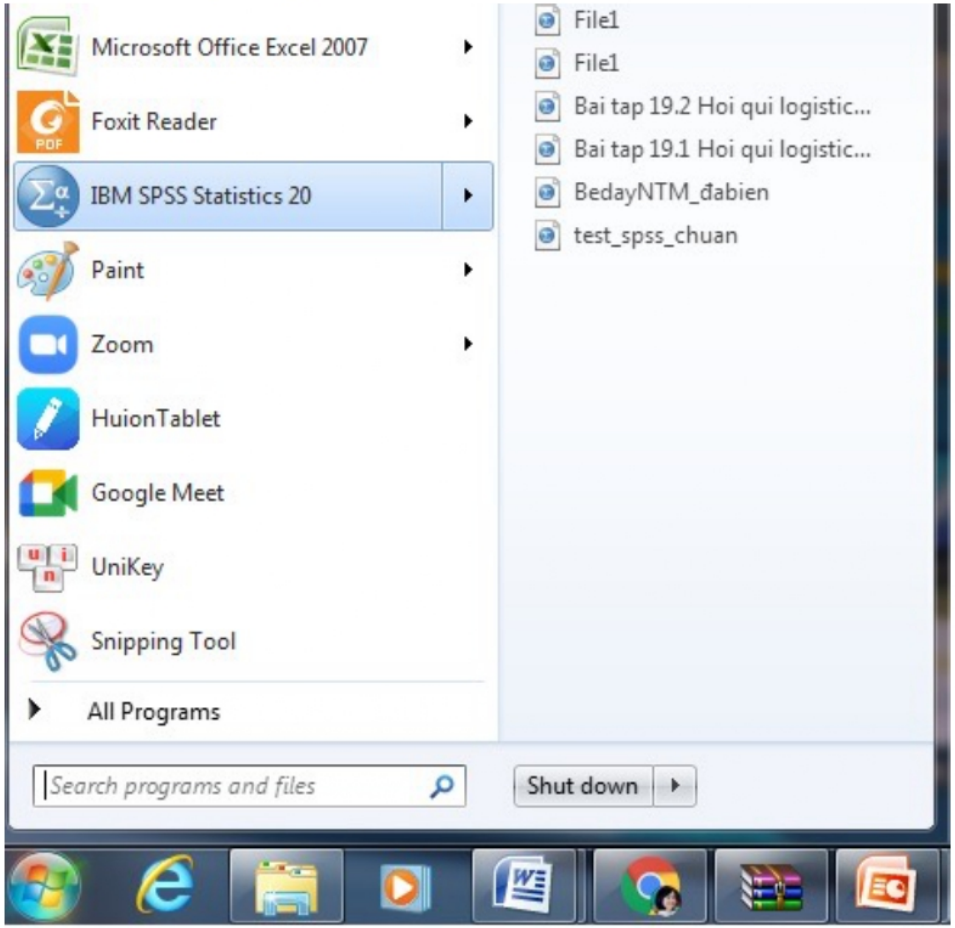
CÀI ĐẶT TIẾNG VIỆT CHO SPSS



Trần Thúy Hiền

Chương 2. Tổng quan về phần mềm SPSS

- Khởi động SPSS: Sử dụng nút Start

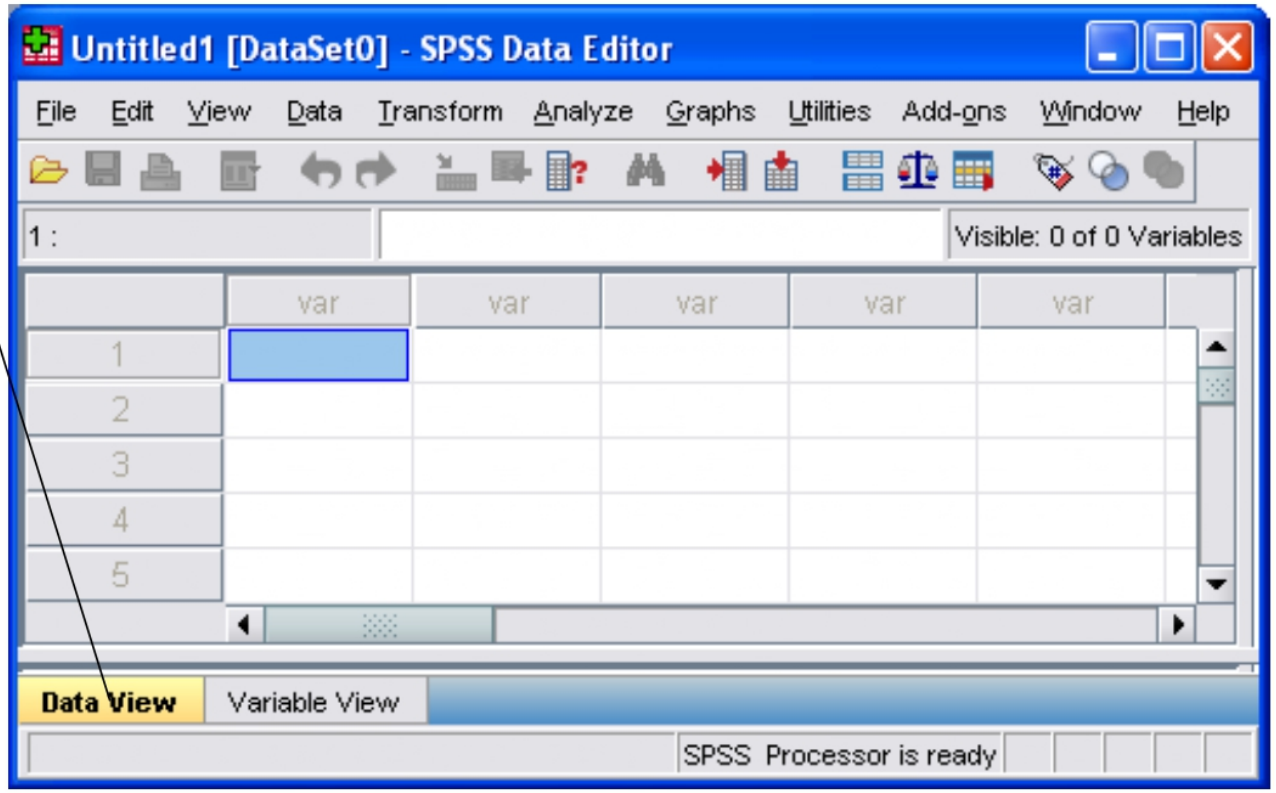


- Tạo Shortcut trên màn hình

Chương 2. Tổng quan về phần mềm SPSS

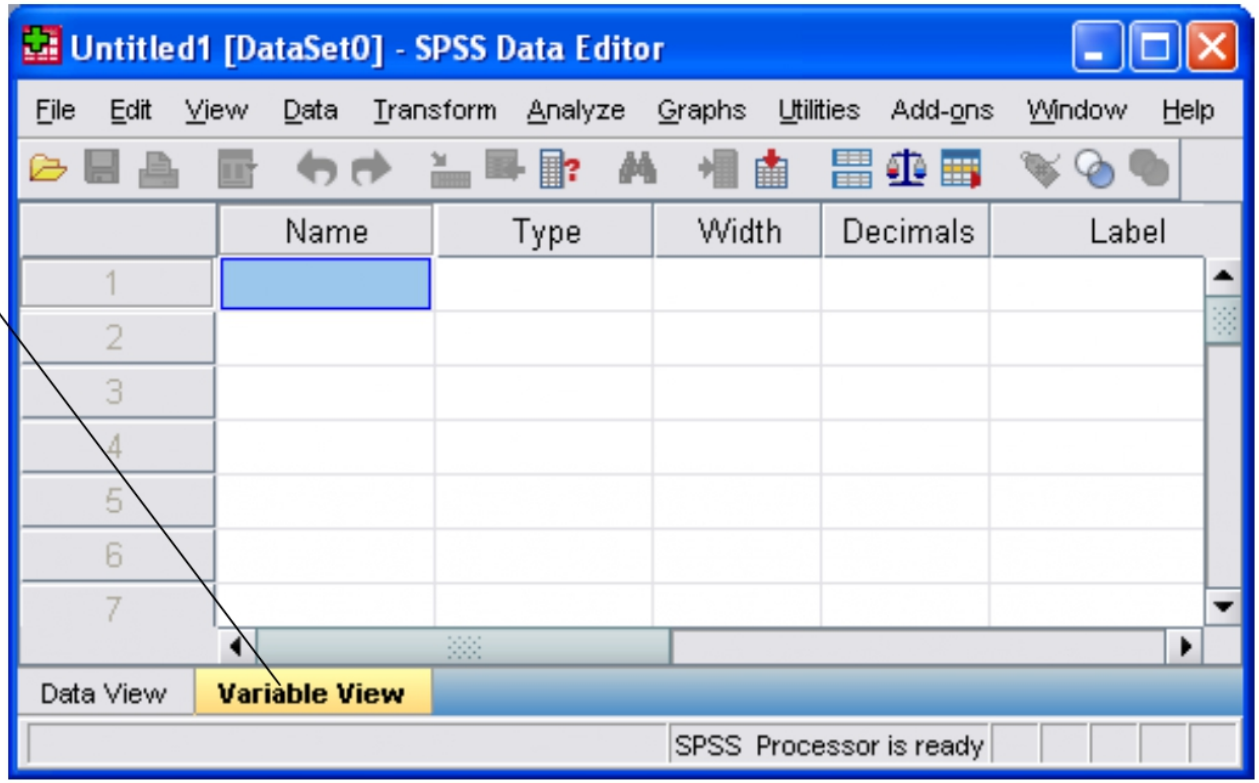
- Giao diện: Cửa sổ chuẩn của SPSS

Cửa sổ nhập liệu (Data View)



Chương 2. Tổng quan về phần mềm SPSS

Cửa số
khai báo biến
(Variable View)

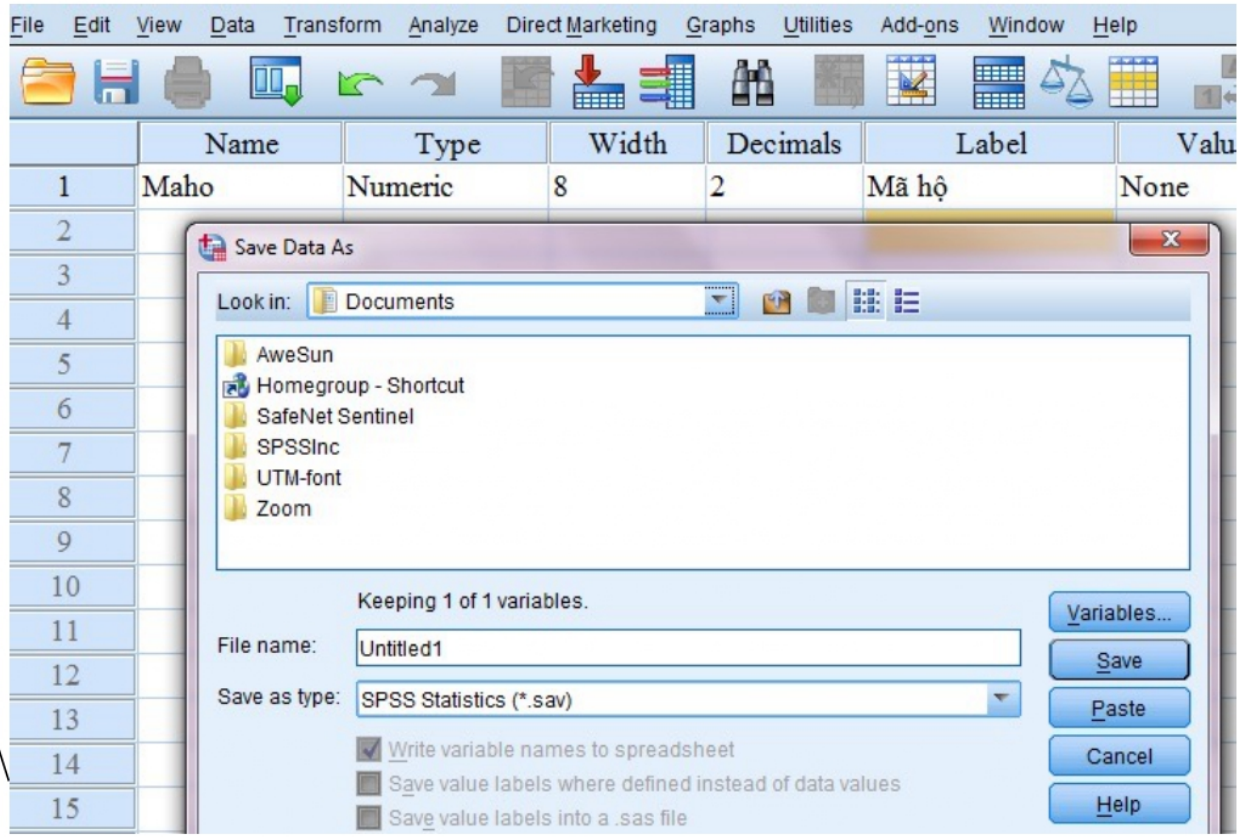


Trần Thúy Hiền

Chương 2. Tổng quan về phần mềm SPSS

File dữ liệu tạo ra cần được lưu vào thư mục thích hợp, có phần mở rộng *. SAV

Lưu File
Save Data As

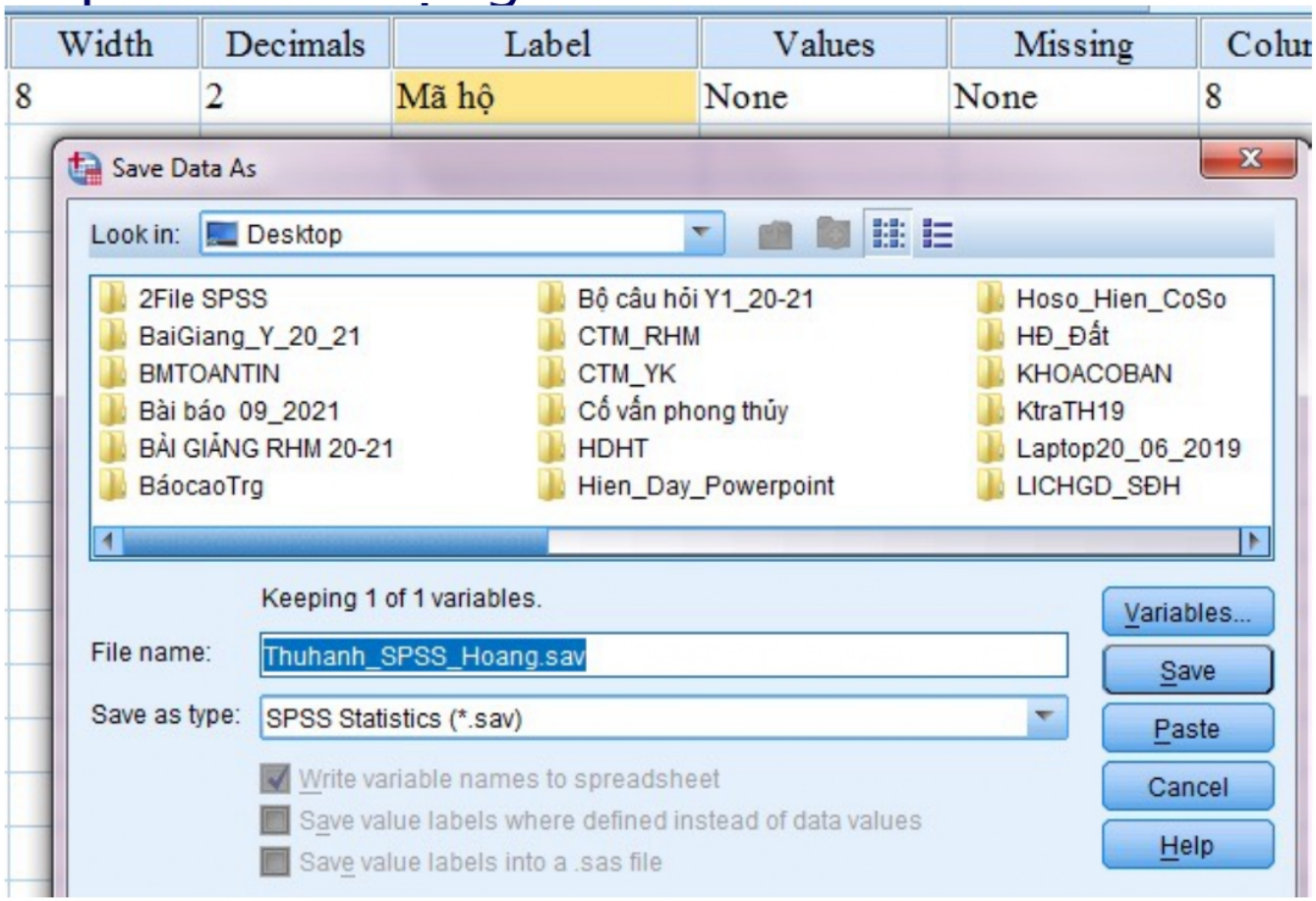


Trần Thúy Hiền

Chương 2. Tổng quan về phần mềm SPSS

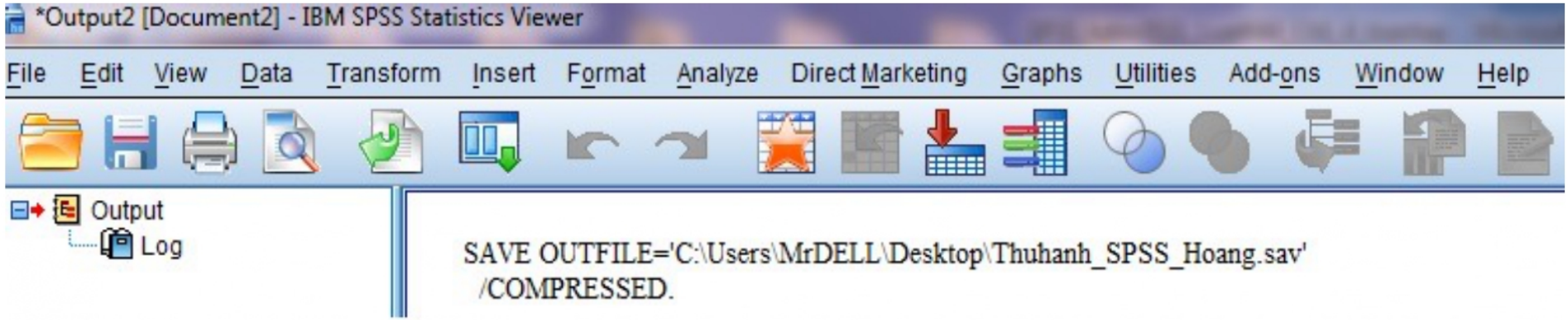
File dữ liệu tạo ra cần được lưu vào thư mục thích hợp, có phần mở rộng *. SAV

Đặt tên File
File Name



Chương 2. Tổng quan về phần mềm SPSS

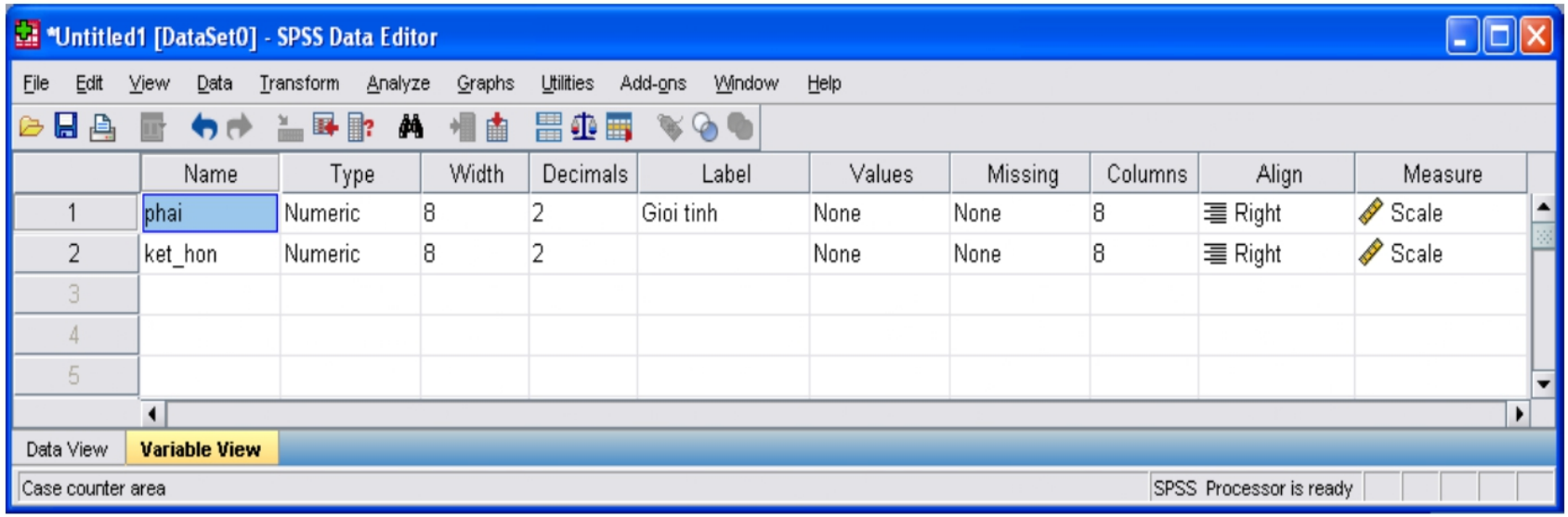
Cửa số Kết quả
(Output)



Trần Thúy Hiền

Chương 2. Tổng quan về phần mềm SPSS

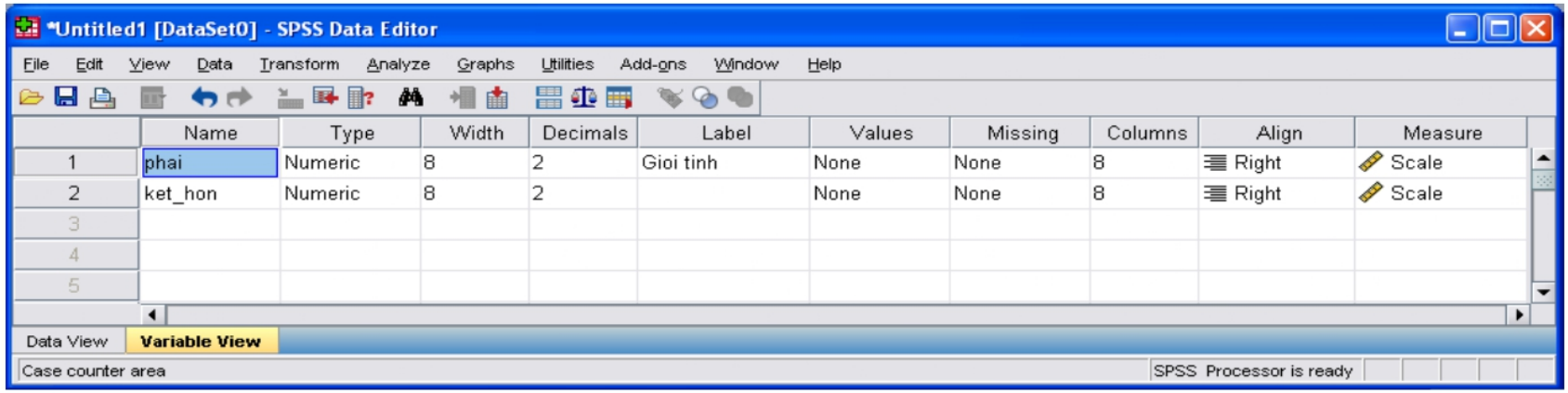
2.3. Khởi tạo biến trên SPSS



- Nhập vào cột Name các biến: phai, ket_hon
- SPSS mặc định dữ liệu có dạng số (Numeric)

Trần Thúy Hiền

Chương 2. Tổng quan về phần mềm SPSS



- **Name**: tên biến **Type**: Kiểu dữ liệu
- **Width**: độ rộng **Decimal**: số chữ số thập phân
- **Label**: nhãn biến **Values**: mã hóa dữ liệu
- **Missing**: khai báo các giá trị khuyết
- **Measure**: thang đo dữ liệu của biến: nominal, ordinal, scale (interval/ratio)

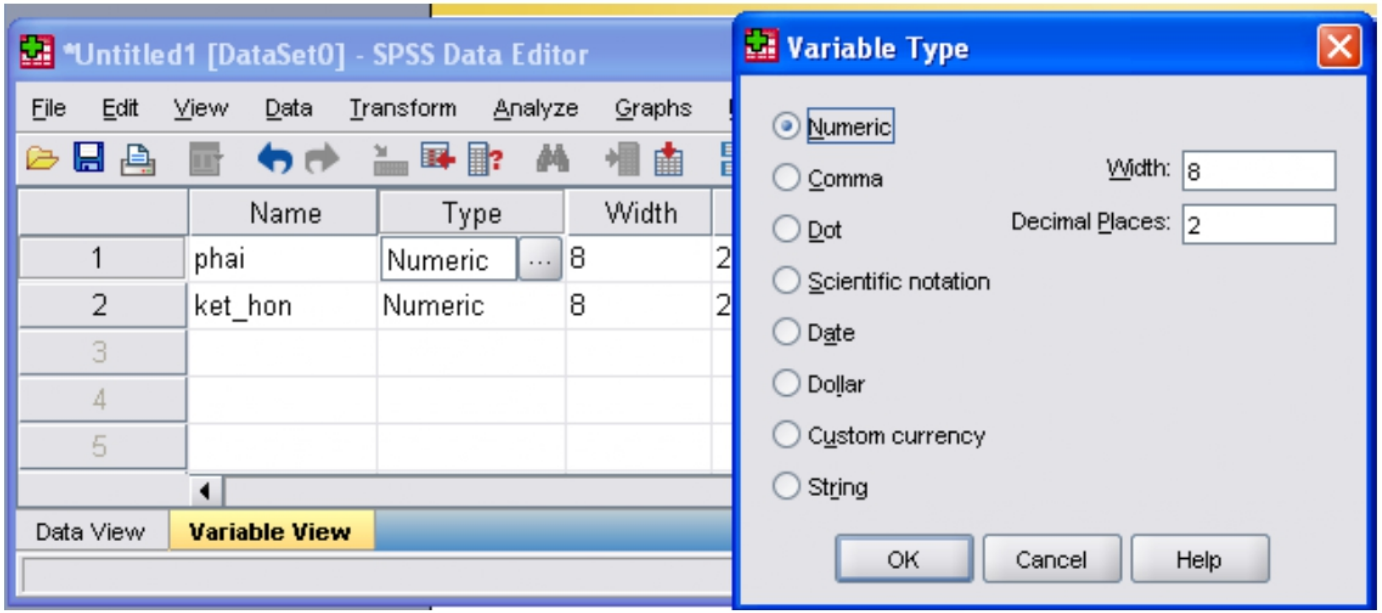
Trần Thúy Hiền

Chương 2. Tổng quan về phần mềm SPSS

- **Name**: tên biến
  - Duy nhất;
  - Phải bắt đầu bằng 1 ký tự;
  - Không chứa khoảng trắng và các ký tự đặc biệt;
- **Width**: độ rộng của biến
  - Là số chữ số hay số ký tự tối đa có thể nhập
- **Decimals**: số chữ số ở phần thập phân (nếu có)
- **Label**: nhãn, tiêu đề của biến
  - Cần đặt ngắn gọn nhưng có tính giải thích cao

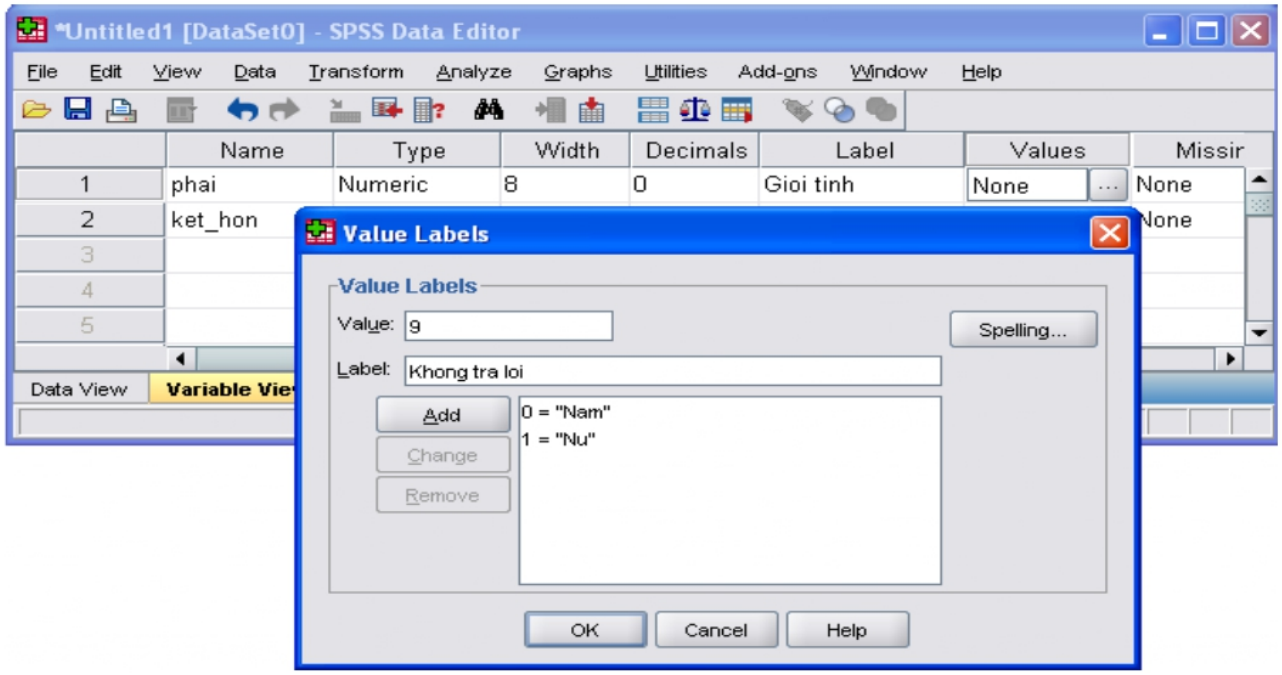
Chương 2. Tổng quan về phần mềm SPSS

- **Type**: kiểu dữ liệu của biến
  - Mặc định là kiểu số;
  - Click vào nút … trong ô Type để thay đổi kiểu dữ liệu của biến



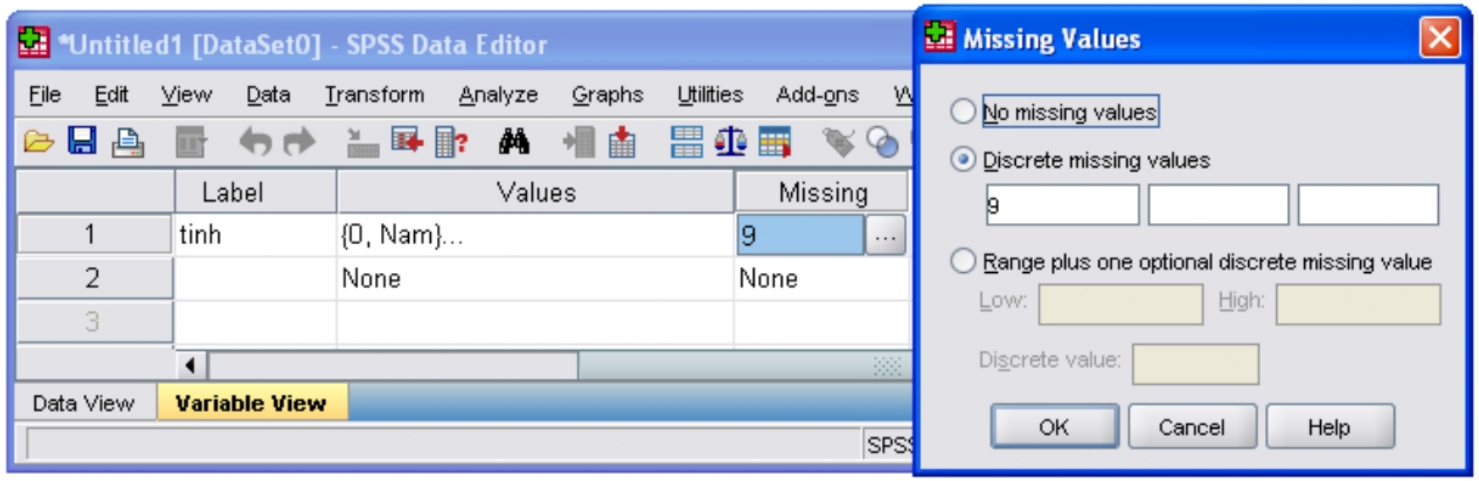
Chương 2. Tổng quan về phần mềm SPSS

- **Values**: mã hóa các biến có thang đo định tính (nominal, ordinal)
  - Click vào nút … ở ô Value để khai báo các chữ số mã hóa



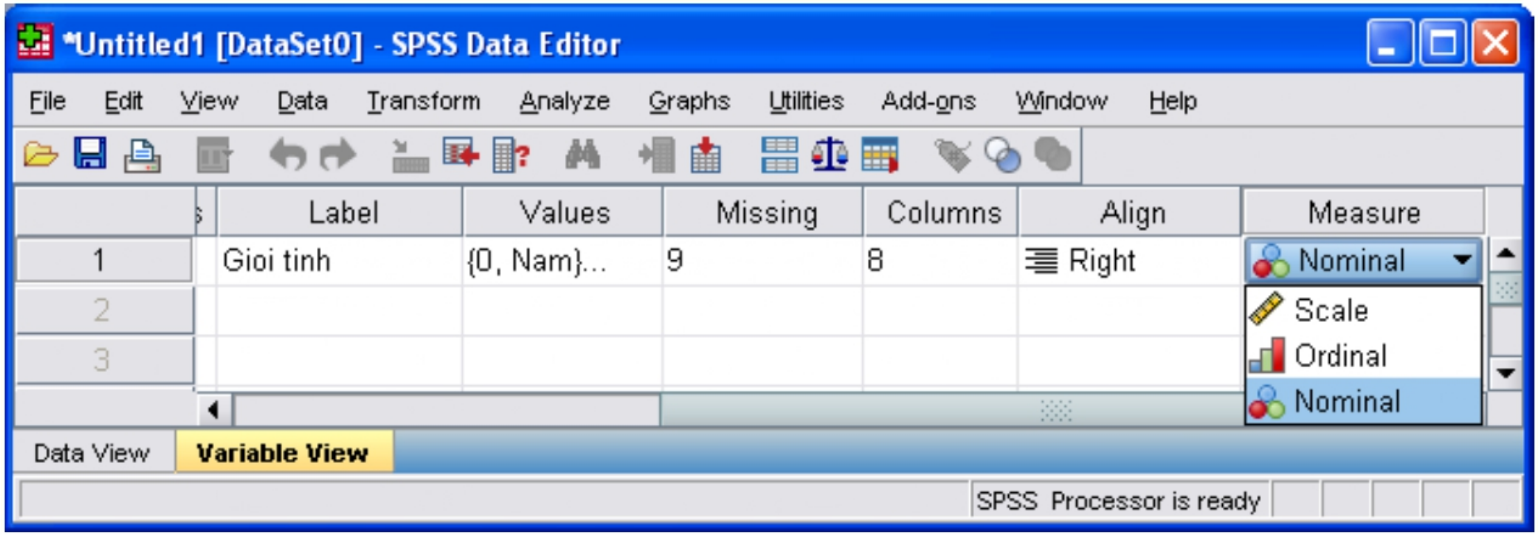
Chương 2. Tổng quan về phần mềm SPSS

- **Missing**: khai báo các loại giá trị khuyết
  - Quy ước giá trị khuyết trước trong Value labels;
  - Khai báo giá trị khuyết đã quy ước trong cửa sổ Missing Values



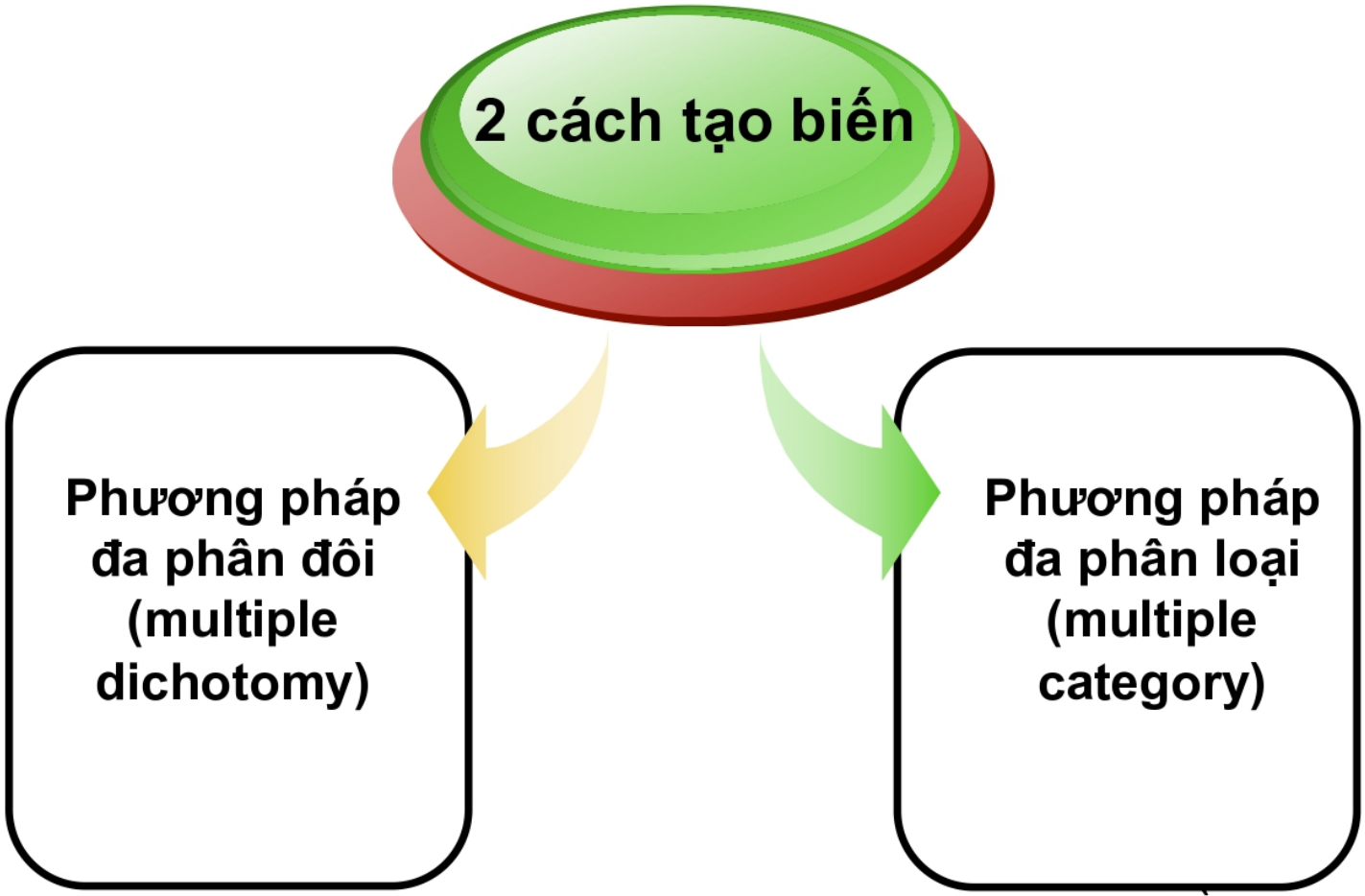
Chương 2. Tổng quan về phần mềm SPSS

- **Measure**: khai báo loại thang đo của biến
  - Khai báo Scale được dùng chung cho dạng thang đo khoảng và thang đo tỷ lệ



Chương 2. Tổng quan về phần mềm SPSS

Mã hóa biến cho câu hỏi nhiều trả lời



Chương 2. Tổng quan về phần mềm SPSS

(1) Phương pháp đa phân đôi (dichotomy)

- Ví dụ: Hãy cho biết anh/chị có thể sử dụng phần mềm nào sau đây:

Microsoft Word	☐	Microsoft Excel	☐	Microsoft PowerPoint	☐
Microsoft Outlook	☐	Microsoft Publisher	☐	Microsoft Project	☐
Microsoft Access	☐	Netscape Navigator	☐	Internet Explorer	☐
Adobe Reader	☐	EndNote	☐	SPSS	☐

Chương 2. Tổng quan về phần mềm SPSS

- Cách giải quyết:
  - Với mỗi trả lời (phần mềm), khai báo một biến; ví dụ: Pmem_word cho Microsoft Word, …
  - Giá trị các biến này được mã hóa theo cách:

    0 = Không biết sử dụng;

    1 = Có thể sử dụng;

Chương 2. Tổng quan về phần mềm SPSS

Id	Pmem_Word	Pmem_Excel	Pmem_Ppt	Pmem_Outlk	Pmem_Pub
1	1	1	1	0	0
2	1	1	1	1	1
3	0	0	0	0	0
4	1	1	1	1	1
5	1	1	1	0	0
6	1	1	1	1	0
7	1	1	1	1	1
8	1	0	0	0	0
9	1	1	1	0	0
10	1	0	0	0	0

Trần Thúy Hiền

Chương 2. Tổng quan về phần mềm SPSS

(2) Phương pháp đa phân loại (category)

- Ví dụ: 6 kỹ năng nào mà anh/chị nghĩ là quan trọng đối với giáo viên dạy môn tin học?

1. _____ 2. _____ 3. _____

4. _____ 5. _____ 6. _____

Với ví dụ này:

- Mỗi người trả lời có thể đưa ra rất nhiều câu trả lời khác nhau;
- Nhưng mỗi người đều có không quá 6 trả lời, nên chúng ta tạo ra 6 biến có tên lần lượt là quality1, …, quality6.

Chương 2. Tổng quan về phần mềm SPSS

- Mã hóa các trả lời có thể có của những người được hỏi như sau:
  - 1 = Khả năng cung cấp các ví dụ thực hành
  - 2 = Khả năng trả lời câu hỏi chính xác
  - 3 = Khả năng giải thích khái niệm rõ ràng
  - 4 = Giảng dạy với tốc độ thích hợp
  - 5 = Hiểu biết về phần mềm
  - 6 = Có tính hài hước
  - 7 = Khác

Chương 2. Tổng quan về phần mềm SPSS

- Tiếp đó, nhập câu trả lời của từng người theo giá trị đã mã hóa;

  Ví dụ: một người chỉ trả lời 3 kỹ năng là: "các ví dụ thực hành", "tính hài hước", và "tốc độ", còn lại để trống -> nhập 1, 6 và 4 cho các biến quality1, quality2, và quality3; các biến còn lại để trống.

Chương 2. Tổng quan về phần mềm SPSS

	quality1	quality2	quality3	quality4	quality5	quality6
1	1	6	4	.	.	.
2	3	2	1	6	7	.
3	2	4	3	6	.	.
4	5	2	3	.	.	.
5	6	2	1	4	.	.
6	3	2	7	.	.	.
7	3	1	.	.	.	.
8	6	3	4	2	.	.
9	5	6	3	.	.	.
10	4	5	6	.	.	.
11	2	1	5	.	.	.
12	6	4	.	.	.	.
13	6	2	4	.	.	.
14	2	6	5	.	.	.
15	7	3	6	2	4	1

Trần Thúy Hiền

Chương 2. Tổng quan về phần mềm SPSS

- **Tóm lại**: trong quá trình khởi tạo biến cần nhận dạng:
  - Loại câu hỏi: một câu trả lời hoặc nhiều câu trả lời
  - Mã hóa biến: chuyển từ dạng chữ sang dạng số
  - Thang đo của biến: nominal, ordinal, scale

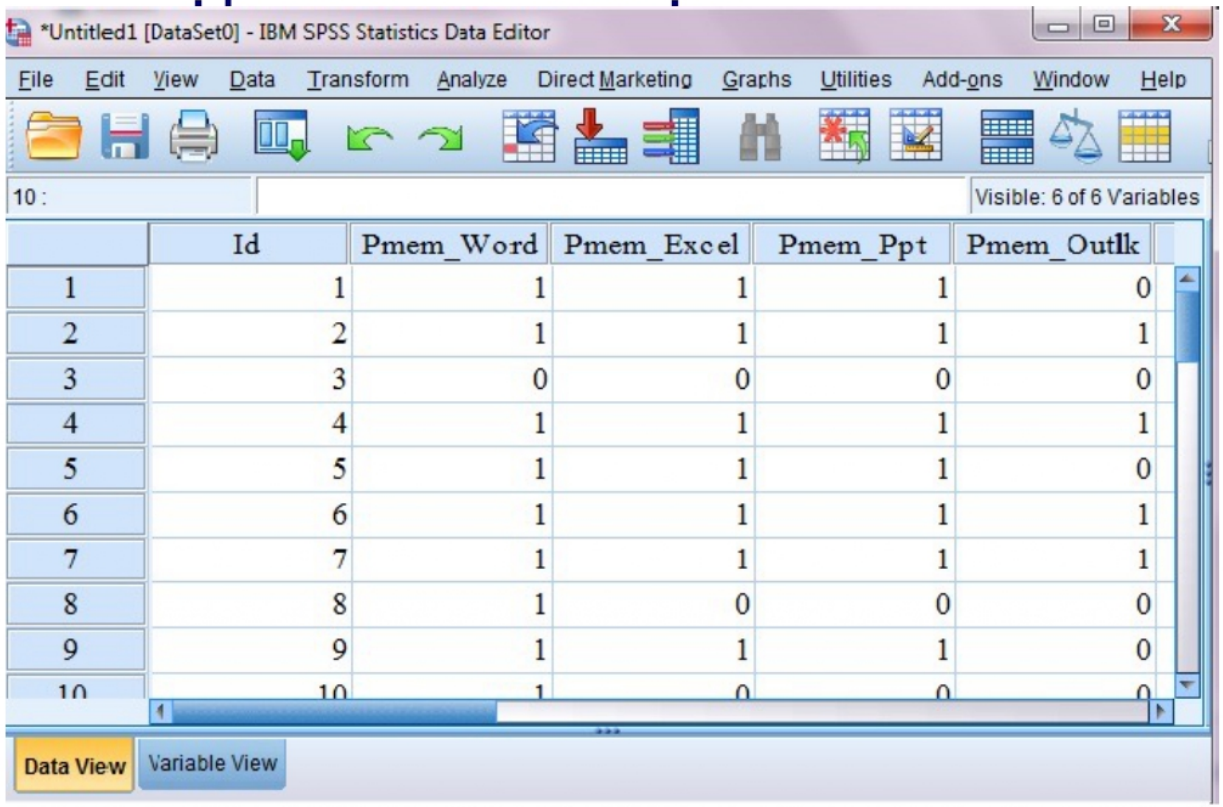
Chương 2. Tổng quan về phần mềm SPSS

- Với câu hỏi chỉ chọn ***một*** trả lời, ta chỉ cần tạo ***một*** biến
- Với câu hỏi có thể chọn ***nhiều*** trả lời, cần phải có ***nhiều*** biến
- Với các dữ liệu **định tính** không phải ký số như "nữ" hay "nhân viên văn phòng" cần phải được tạo một con số mã hóa.
  - Ví dụ: Nam nhập "1", nữ nhập "2"
- Các thông tin thu thập dưới dạng dữ liệu **định lượng** đã ở dưới dạng số nên không cần mã hóa.

2.4. Nhập dữ liệu và làm sạch dữ liệu

(1) Nhập dữ liệu

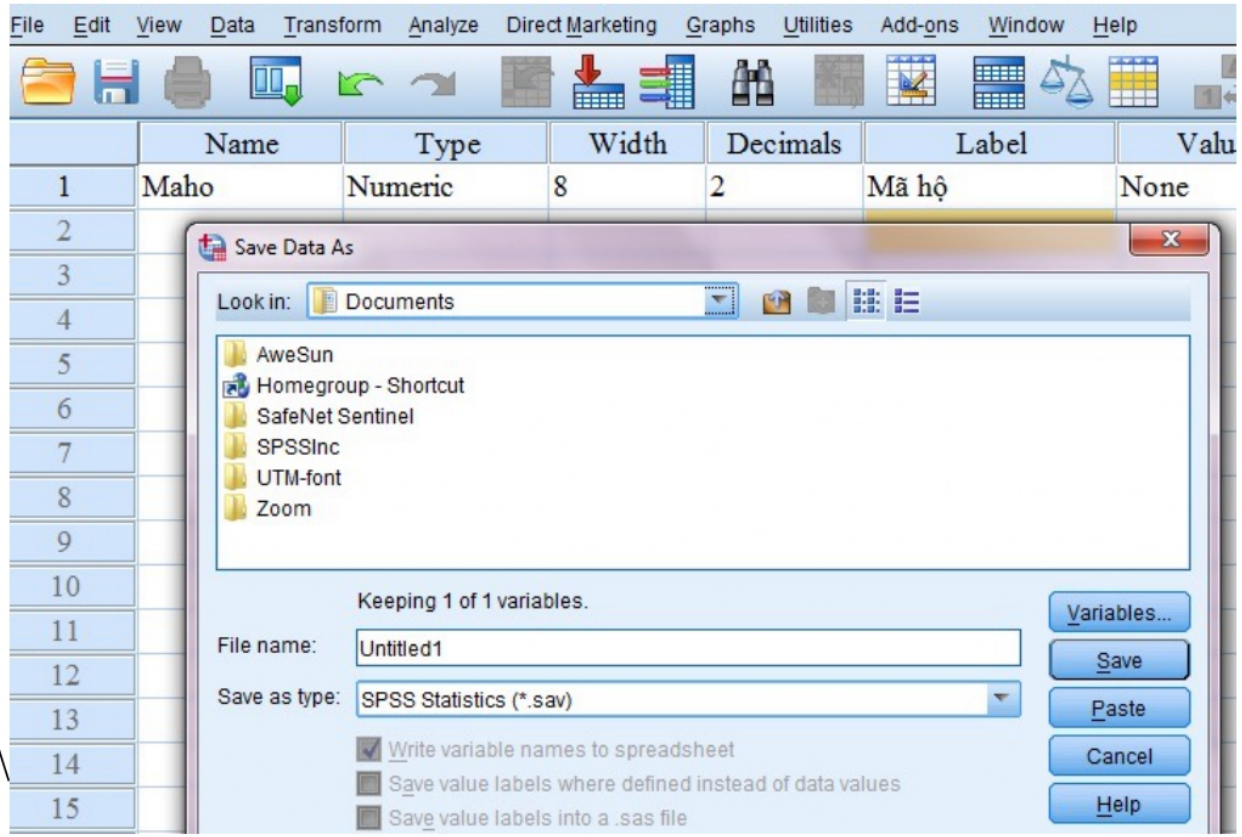
- Data View: Chọn ô cần nhập dữ liệu; Gõ giá trị cần nhập vào ô đã chọn



Chương 2. Tổng quan về phần mềm SPSS

File dữ liệu tạo ra cần được lưu vào thư mục thích hợp, có phần mở rộng *. SAV

Lưu File
Save Data As

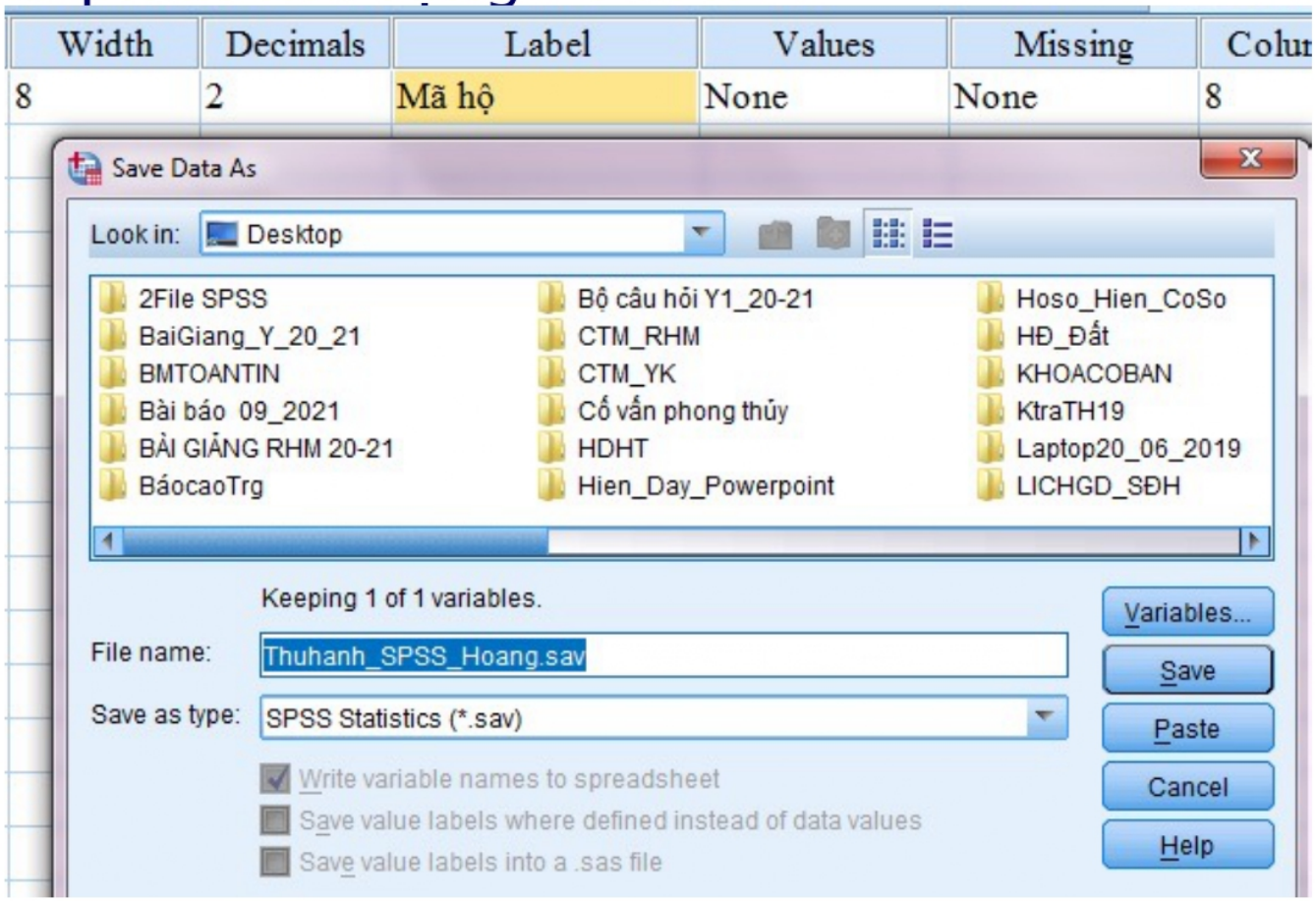


Trần Thúy Hiền

Chương 2. Tổng quan về phần mềm SPSS

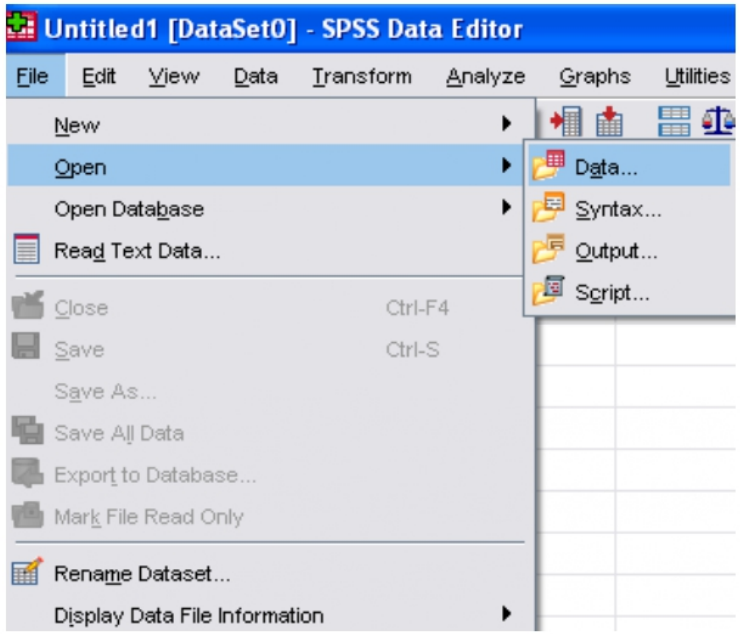
File dữ liệu tạo ra cần được lưu vào thư mục thích hợp, có phần mở rộng *. SAV

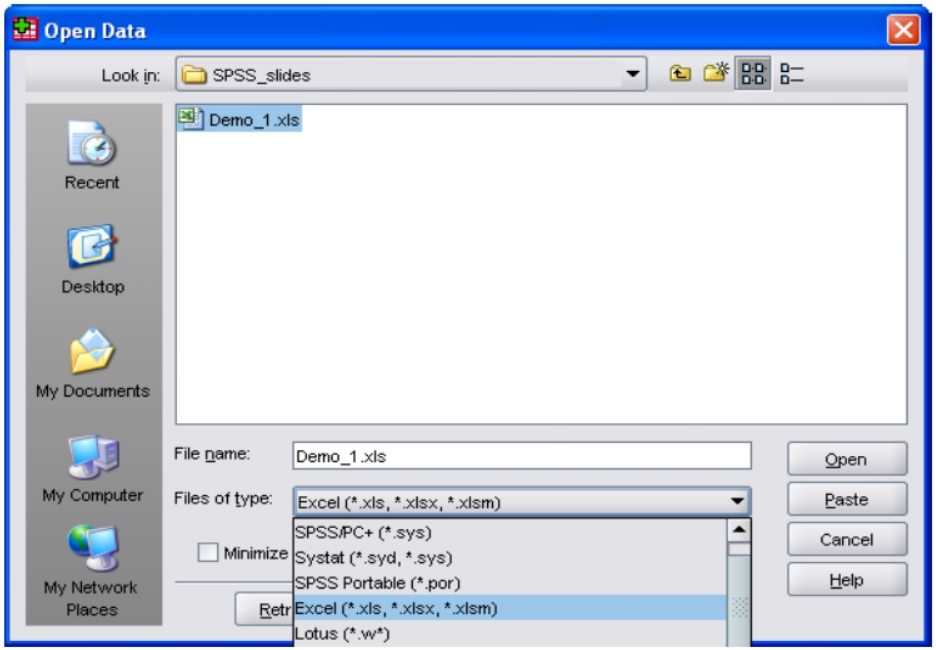
Đặt tên File
File Name



Chương 2. Tổng quan về phần mềm SPSS

- Đọc dữ liệu từ file Excel: Trong cửa sổ Open Data, phần **Files of type** chọn kiểu file Excel (*.xls).

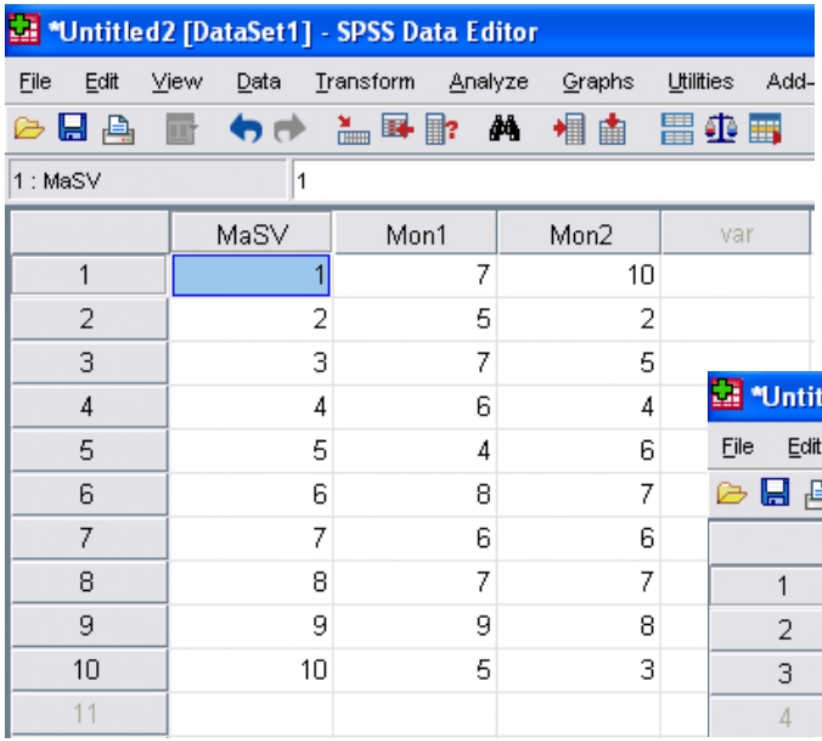




Trần Thúy Hiền

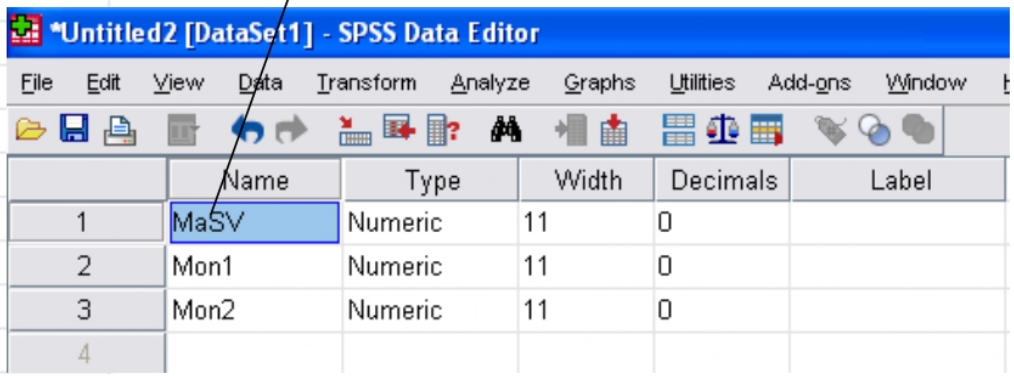
Chương 2. Tổng quan về phần mềm SPSS

- Kết quả



	MaSV	Mon1	Mon2	var
1	1	7	10	
2	2	5	2	
3	3	7	5	
4	4	6	4	
5	5	4	6	
6	6	8	7	
7	7	6	6	
8	8	7	7	
9	9	9	8	
10	10	5	3	
11				

Vào cửa sổ Variable View để chỉnh sửa lại một số thuộc tính của các biến



	Name	Type	Width	Decimals	Label
1	MaSV	Numeric	11	0	
2	Mon1	Numeric	11	0	
3	Mon2	Numeric	11	0	
4					

Chương 2. Tổng quan về phần mềm SPSS

(2) Làm sạch dữ liệu

- Dữ liệu đơn giản
  - **Data/Sort Cases…**
  - Chọn biến cần sắp xếp để kiểm tra
- Dữ liệu phức tạp
  - Data/Select Cases…
  - Chọn If condition is satisfied, click nút If… để xây dựng biểu thức điều kiện.
- Sử dụng bảng chéo Crosstab để tìm lỗi
  - Analyze/Descriptive Statistics/Crosstabs…
  - Chọn các biến cần xem xét sự liên quan (row, column)

Thank You !
Email: tthien@huemed-univ.edu.vn

Chương 3. Các thao tác cơ bản trên SPSS

Data

3.1. Lựa chọn các quan sát (Select cases)

3.2. Tách tập dữ liệu (Split file)

Transform

3.3. Tính toán giữa các biến (Compute)

3.4. Mã hóa lại biến (Recode)

3.5. Công cụ đếm (Count)

Trần Thúy Hiền

3.1. Lựa chọn các quan sát (Select cases)

1. Chức năng

- Chọn ra những nhóm nhỏ các trường hợp quan sát dựa trên tiêu chuẩn hay điều kiện cụ thể;
- Chọn một mẫu ngẫu nhiên các trường hợp quan sát từ tổng thể dữ liệu;
- Các phân tích thống kê sẽ chỉ thực hiện trên các quan sát đã được lựa chọn.

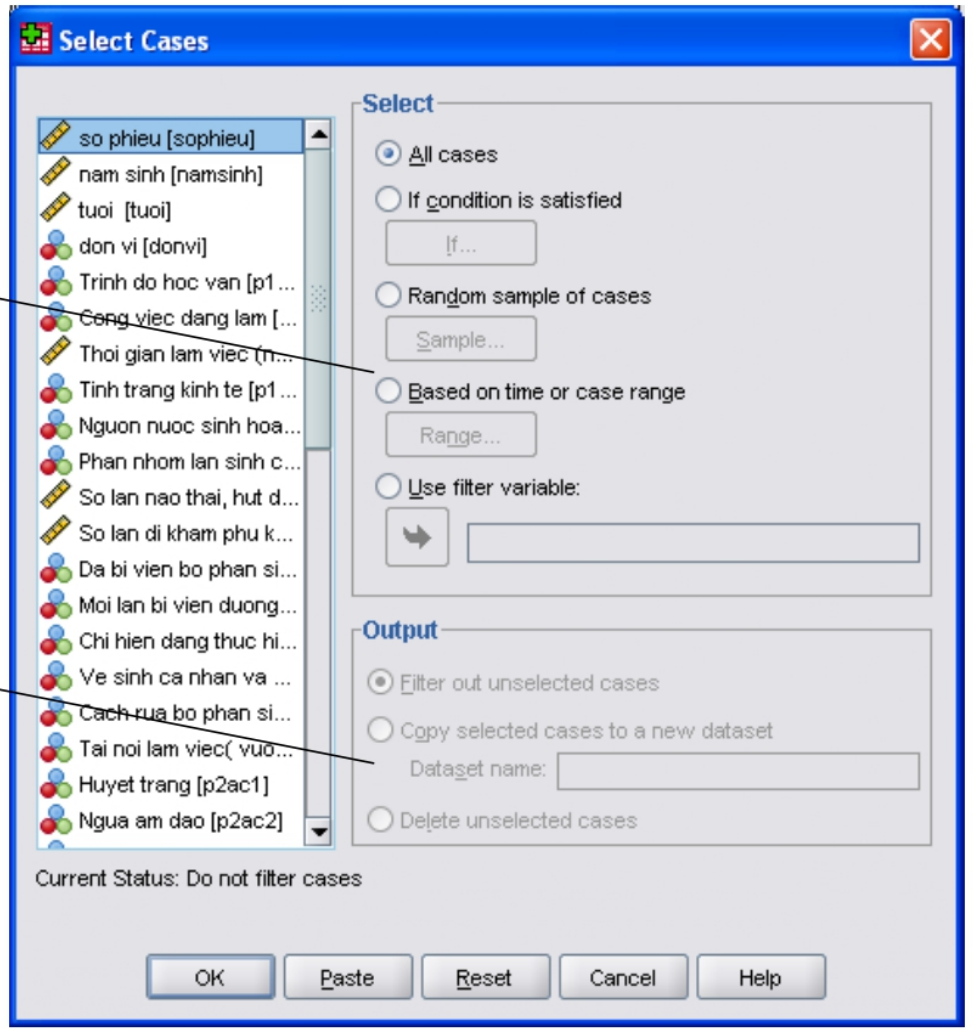
3.1. Lựa chọn các quan sát (Select cases)

2. Thao tác thực hiện

- Data/Select Cases..

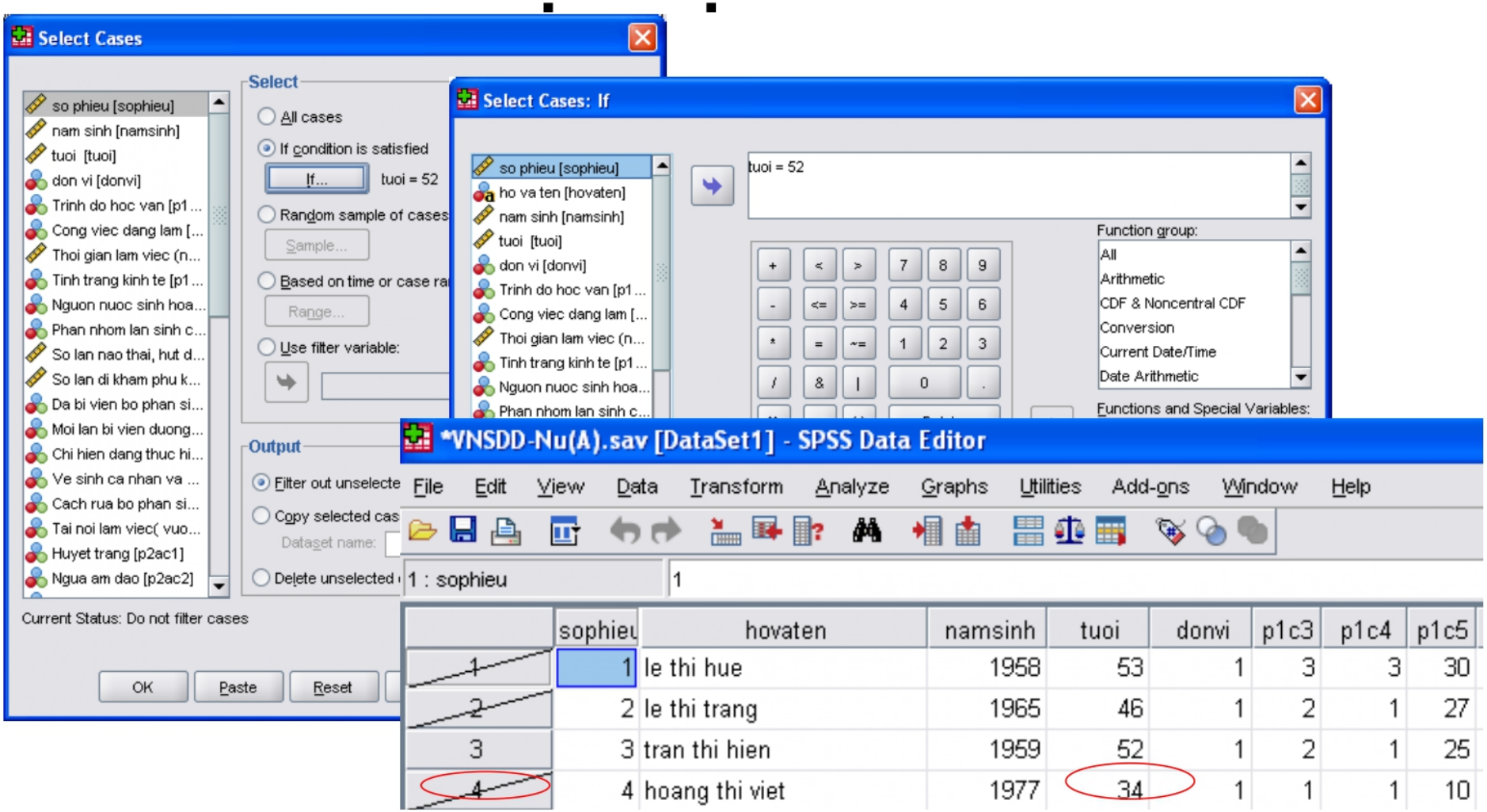
Các cách thức lựa chọn các quan sát

Các cách thức đối xử đối với các quan sát không được lựa chọn



3.1. Lựa chọn các quan sát (Select cases)

2. Thao tác thực hiện



Trần Thúy Hiền

3.2. Tách tập dữ liệu (Split file)

1. Chức năng

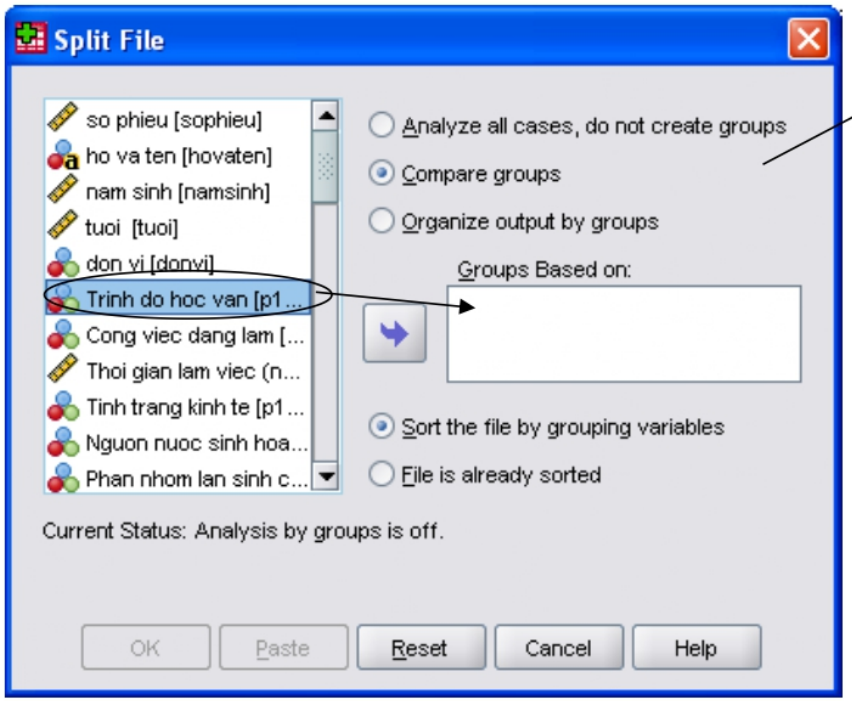
- Tách dữ liệu trong tập dữ liệu đang quan sát thành những nhóm nhỏ riêng biệt
- Các phân tích thống kê sẽ cho các kết quả được thực hiện riêng biệt theo từng nhóm nhỏ dữ liệu đã được tách.

Ví dụ. Tách tập dữ liệu theo biến "trình độ học vấn", "giới tính", "vùng sinh sống"….

3.2. Tách tập dữ liệu (Split file)

2. Thao tác thực hiện

- Data/Split File…



Các hình thức tách hoặc không phân tách dữ liệu

Trần Thúy Hiền

3.3. Tính toán giữa các biến (Compute)

1. Chức năng

- Tính toán giữa các giá trị trong các biến
- Kết quả sẽ được lưu giữ trong:
  - Biến mới
  - Một biến khác sẵn có
  - Biến chứa đựng giá trị đang tính toán

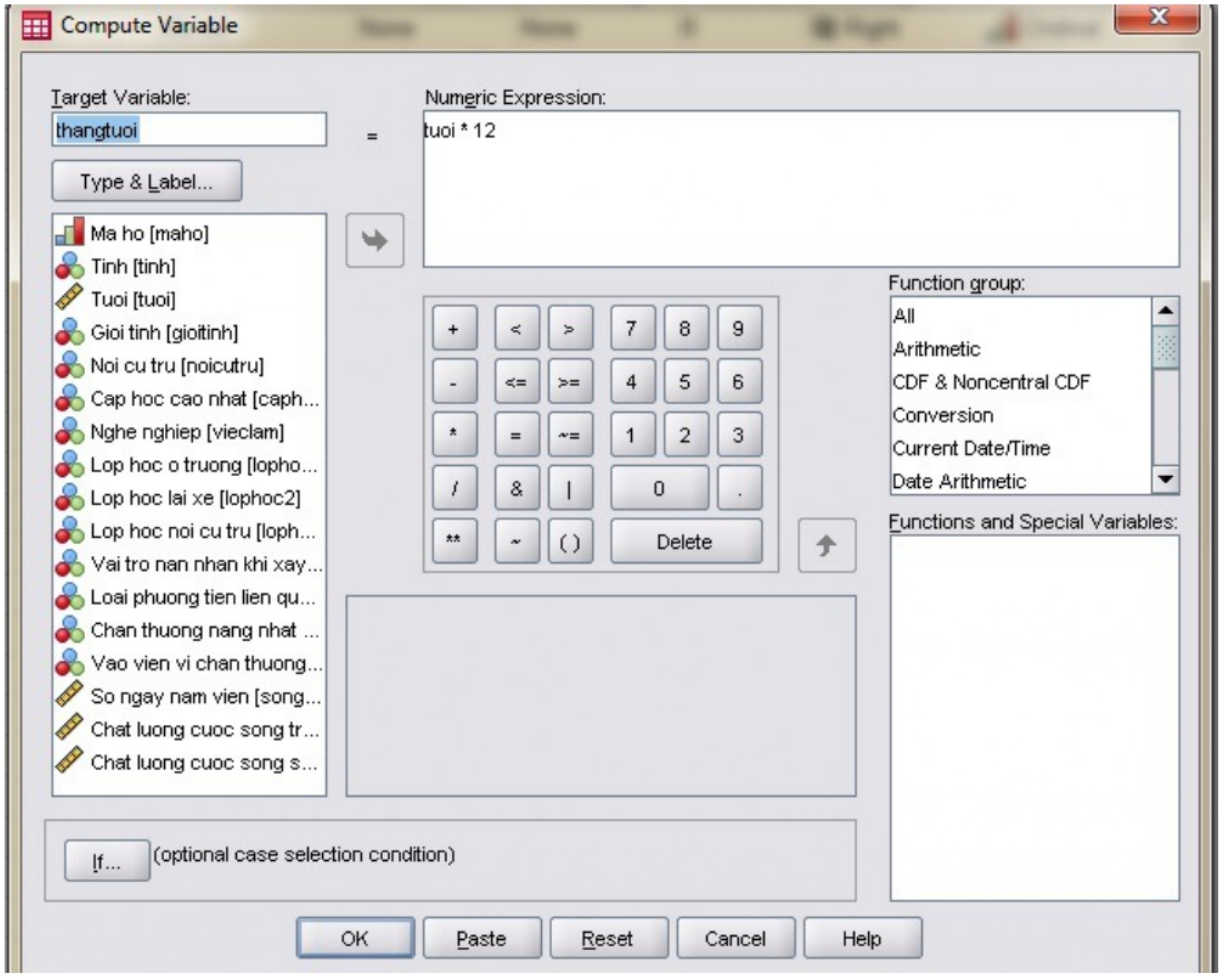
Ví dụ. Tạo biến mới **thangtuoi** theo yêu cầu:

thangtuoi = **tuoi** * 12

3.3. Tính toán giữa các biến (Compute)

2. Thao tác thực hiện

- Transform/Compute Variable…



3.4. Mã hóa lại biến (Recode)

1. Chức năng

- Mã hóa lại các giá trị của biến đã thu thập được thành các giá trị mã hóa mới phù hợp với đòi hỏi của quá trình phân tích dữ liệu
- Có hai loại mã hóa lại biến (recode):
  - Recode trên cùng một biến
  - Recode vào biến khác

    Ví dụ. Mã hóa biến **tuoi** thành một biến mới là **nhomtuoi** theo các nhóm 1-15, 16-22, 23-35, 36-50, 51-70.

3.4. Mã hóa lại biến (Recode)

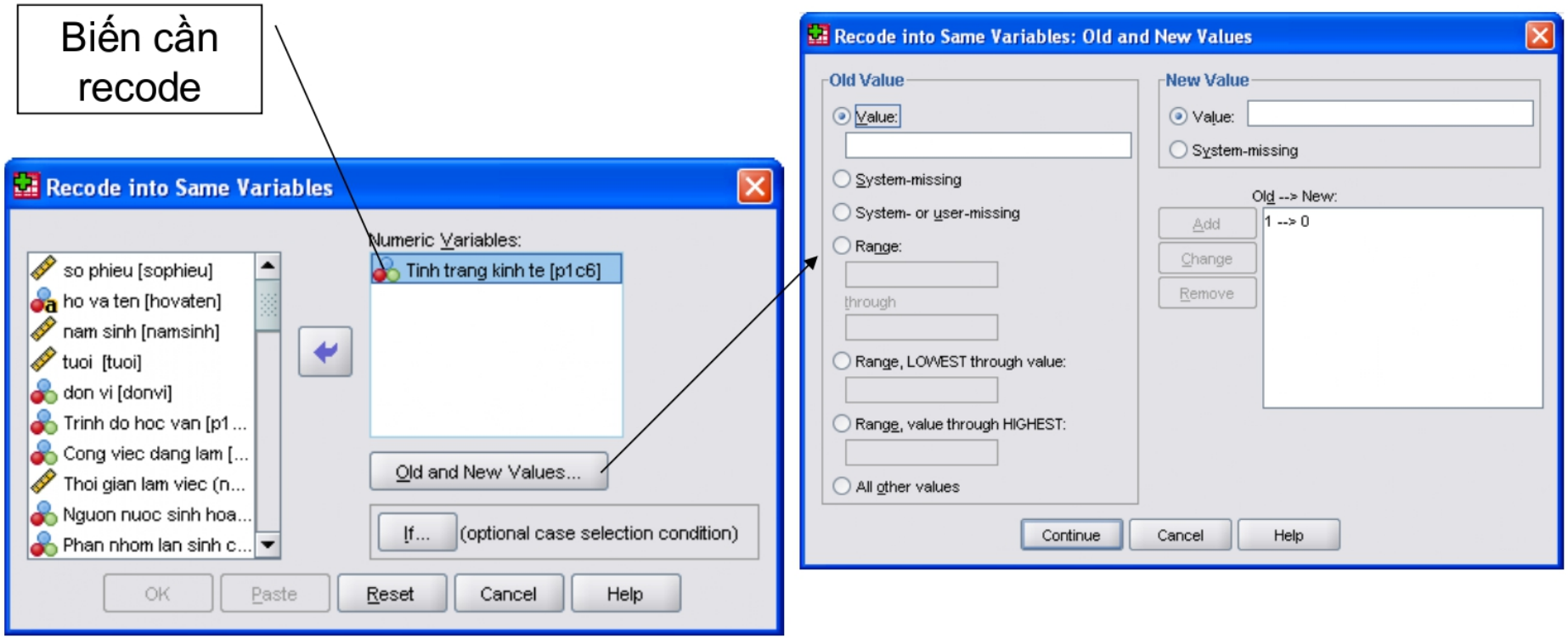
2. Recode trên cùng một biến

- Mã hóa lại những giá trị trong một biến hiện có thành những giá trị mới;
- Các giá trị mới thu được sẽ nằm ngay trong biến hiện hành và thay thế các giá trị cũ của biến đó.

3.4. Mã hóa lại biến

2. Recode trên cùng một biến

- Transform/Recode into Same Variables…



Trần Thúy Hiền

3.4. Mã hóa lại biến (Recode)

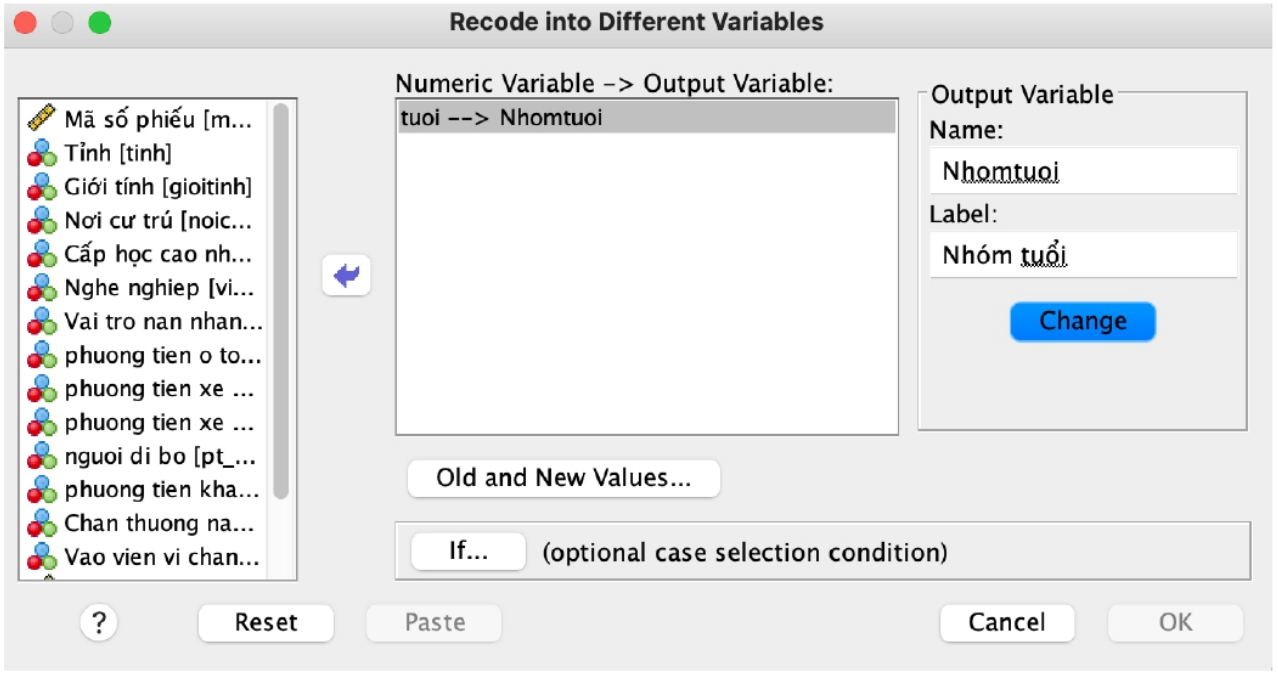
3. Recode vào một biến khác

- Mã hóa lại những giá trị trong một biến hiện có thành những giá trị mới;
- Các giá trị mới thu được sẽ nằm trong một biến mới được tạo ra trong cơ sở dữ liệu.
- Biến cũ vẫn được lưu giữ với các giá trị mã hóa cũ của nó trong cơ sở dữ liệu.

3.4. Mã hóa lại biến (Recode)

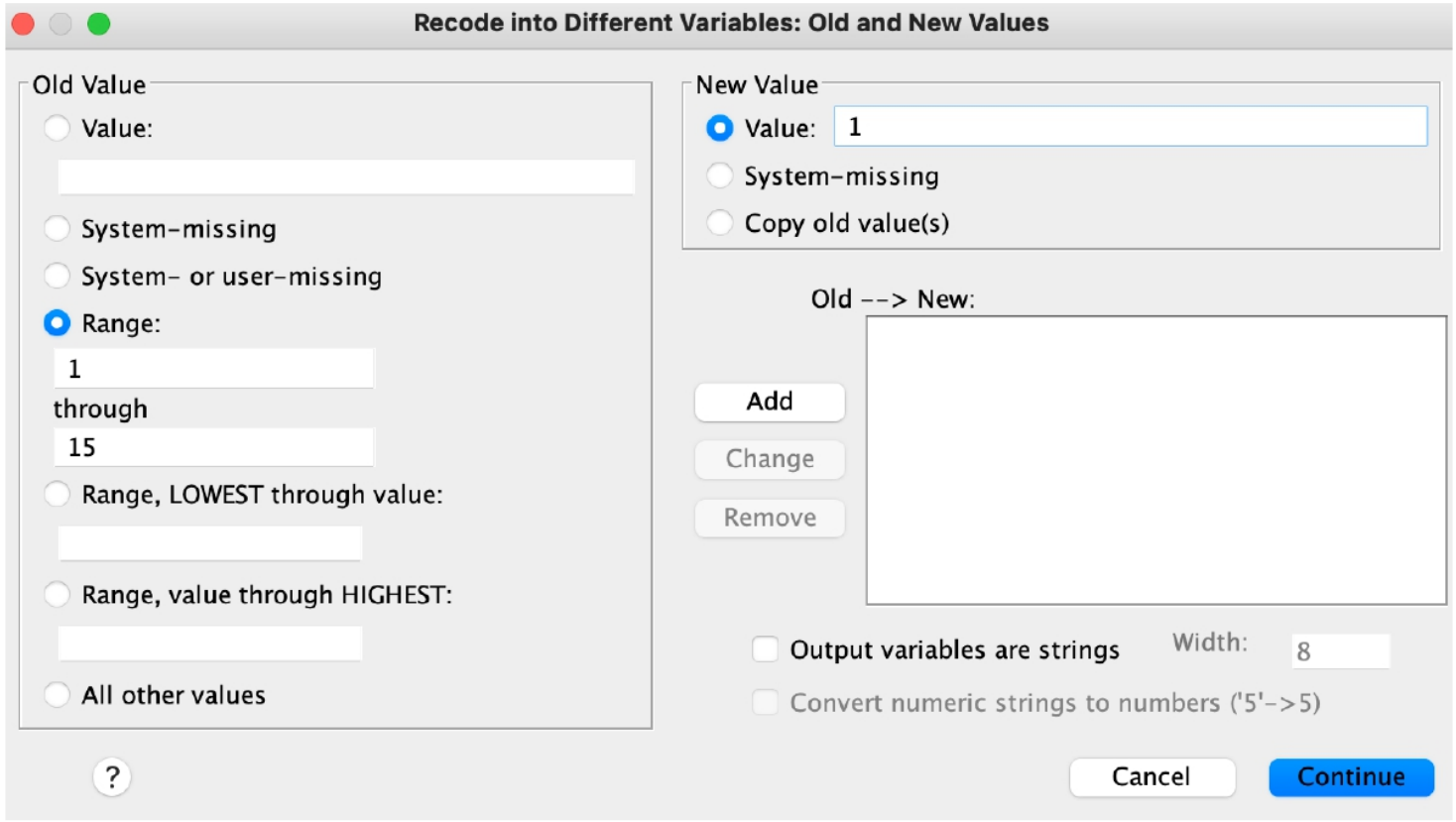
3. Recode vào một biến khác

- Transform/Recode into Different Variables...



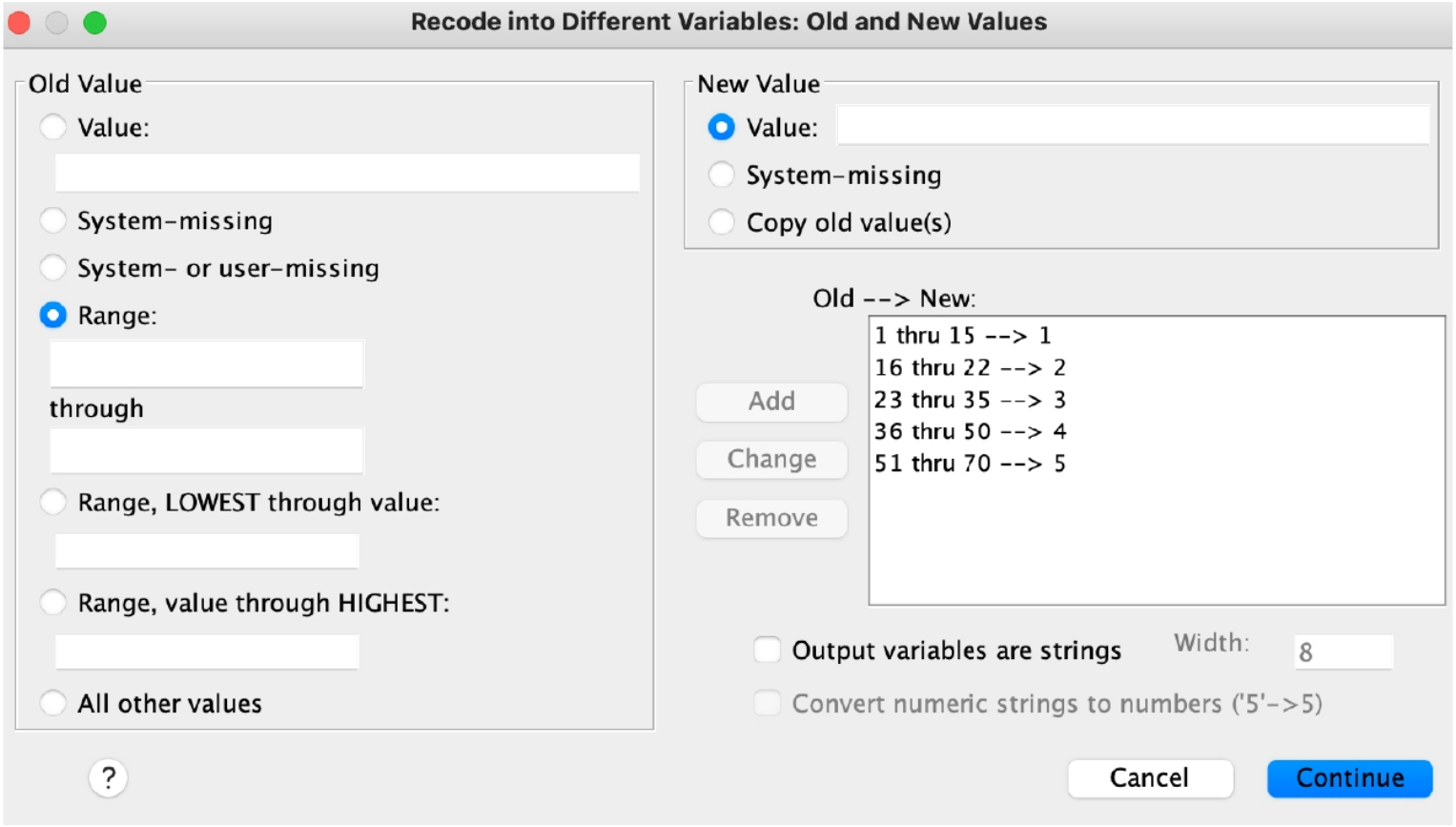
Trần Thúy Hiền

3.4. Mã hóa lại biến (Recode)



Trần Thúy Hiền

3.4. Mã hóa lại biến (Recode)



Trần Thúy Hiền

3.4. Mã hóa lại biến (Recode)

4. Recode tự động

- Mã hóa tự động các giá trị dạng chuỗi sang dạng số vào trong một biến mới
- Biến mới sẽ chứa các số nguyên, mỗi số nguyên đó đại diện cho các giá trị dạng chuỗi giống nhau
- Các giá trị dạng chuỗi được mã hóa theo thứ tự alphabet
- Transform/Automatic Recode…

3.5. Công cụ đếm (Count)

1. Chức năng

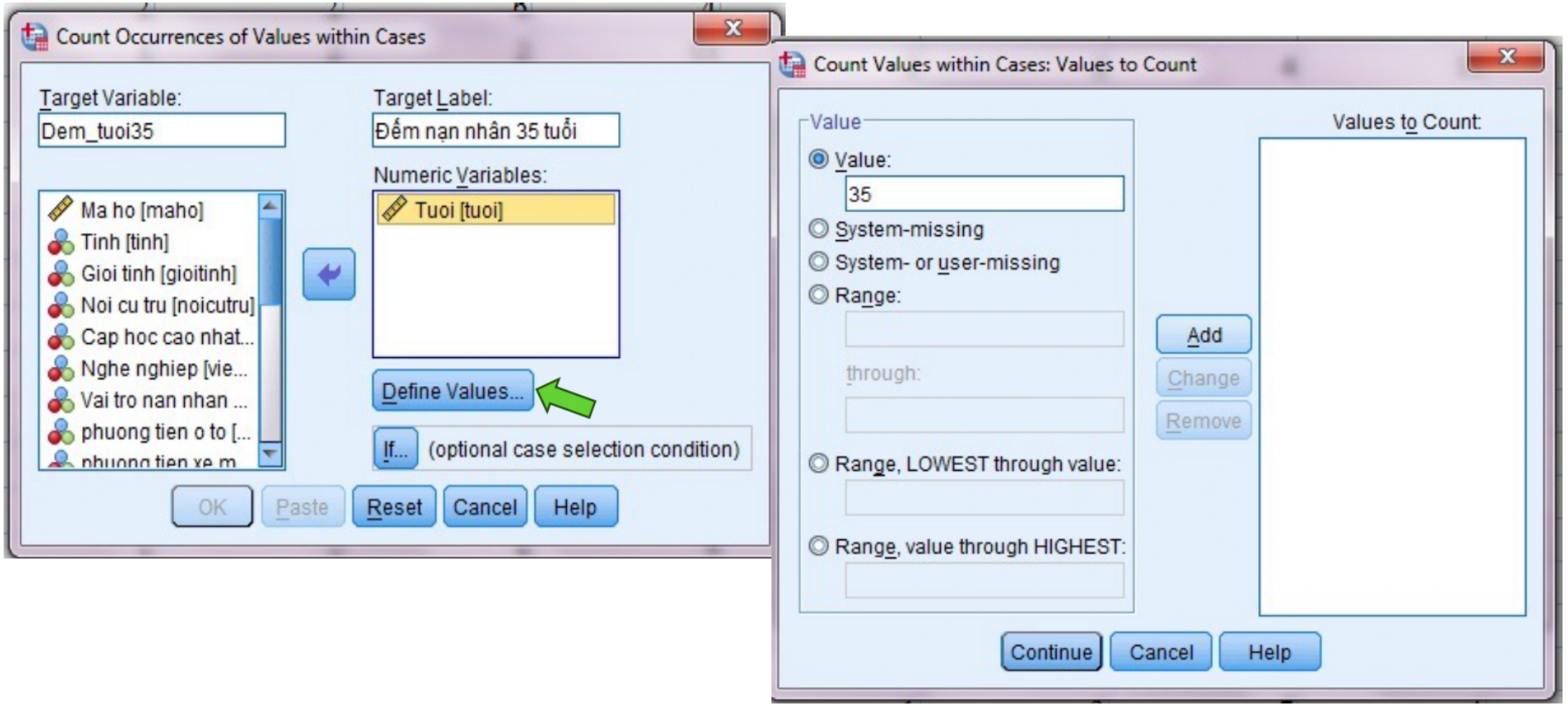
- Tạo ra 1 biến mới chứa kết quả số lần xuất hiện (số đếm) của một hay nhiều giá trị được chỉ định ra trong danh sách các biến được chọn.

Ví dụ. Tạo biến “Dem_tuoi35” đếm (xác định) người trong mẫu có tuổi là 35.

3.5. Công cụ đếm (Count)

2. Thao tác thực hiện

- Transform/Count Values within Cases…



Trần Thúy Hiền

Thank You !
Email: tthien@huemed-univ.edu.vn

Chương 4. SPSS trong Thống kê mô tả

4.1. Lý thuyết thống kê và Một số đại lượng thống kê

4.2. Bảng tần số (Frequencies)

4.3. Bảng thống kê mô tả (Descriptives)

4.4. Khám phá dữ liệu (Explore)

4.5. Bảng chéo (Crosstabs)

4.6. Bảng tùy chỉnh (Custom Tables)

4.7. Xử lý câu hỏi nhiều trả lời

4.8. Biểu diễn dữ liệu bằng biểu đồ

Trần Thúy Hiền

4.1. Lý thuyết thống kê và Một số đại lượng thống kê

4.1.1. Thống kê trong nghiên cứu khoa học

Thống kê có ứng dụng rộng rãi trong mọi lĩnh vực đời sống xã hội, chiếm một vị trí quan trọng trong việc thực hiện các công trình nghiên cứu khoa học:

"Một bài báo khoa học nếu không được phân tích thống kê đúng phương pháp thì sẽ không có giá trị về mặt khoa học"

4.1. Lý thuyết thống kê và Một số đại lượng thống kê

(1) Thống kê

- Thống kê là một lĩnh vực của khoa học giúp người nghiên cứu tổ chức một công cuộc nghiên cứu khoa học để tìm ra phương pháp tổ chức công tác thực tế một cách hợp lý và hiệu quả nhất;
- Thống kê là khoa học liên quan đến việc thu thập, phân tích, diễn giải hay giải thích, trình bày các dữ liệu và dùng chúng để kiểm định các giả thuyết.

4.1. Lý thuyết thống kê và Một số đại lượng thống kê

(2) Đối tượng, mục tiêu

- Đối tượng nghiên cứu của thống kê là quần thể (tổng thể, dân số)
- Mục tiêu của thống kê là suy luận (tiên đoán và quyết định) về quần thể dựa vào các dữ kiện của mẫu rút ra từ quần thể và đại diện cho quần thể đó

4.1. Lý thuyết thống kê và Một số đại lượng thống kê

(3) Các vấn đề cơ bản của thống kê

- Xác định đối tượng, mục tiêu nghiên cứu và chỉ tiêu quan sát
- Phương pháp chọn mẫu
- Thu thập, phân tích các dữ liệu mẫu
- Suy đoán về quần thể dựa vào mẫu

4.1. Lý thuyết thống kê và Một số đại lượng thống kê

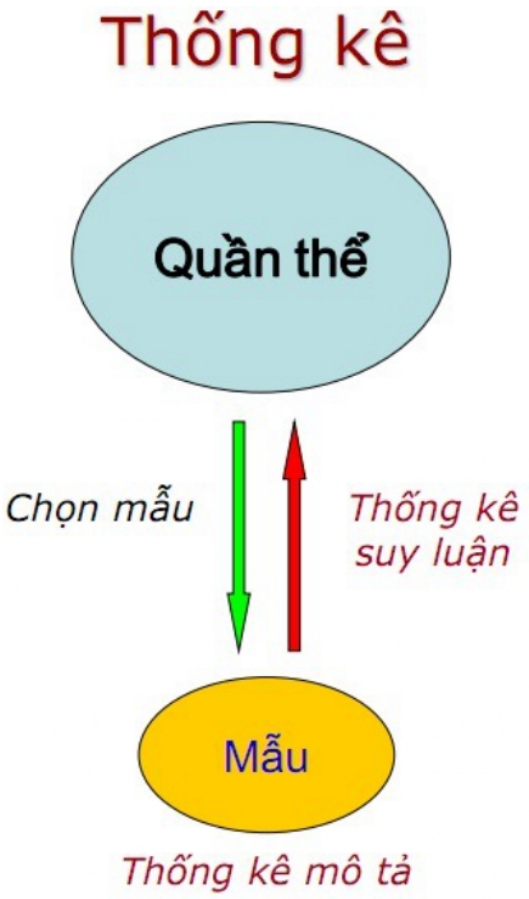
- *Thống kê mô tả*:

  là phương pháp dùng để tóm tắt hoặc mô tả một tập hợp dữ liệu.

- *Thống kê suy luận*:

  là phương pháp mô hình hóa trên các dữ liệu quan sát, giải thích các biến thiên có tính ngẫu nhiên và tính không chắc chắn của các quan sát.

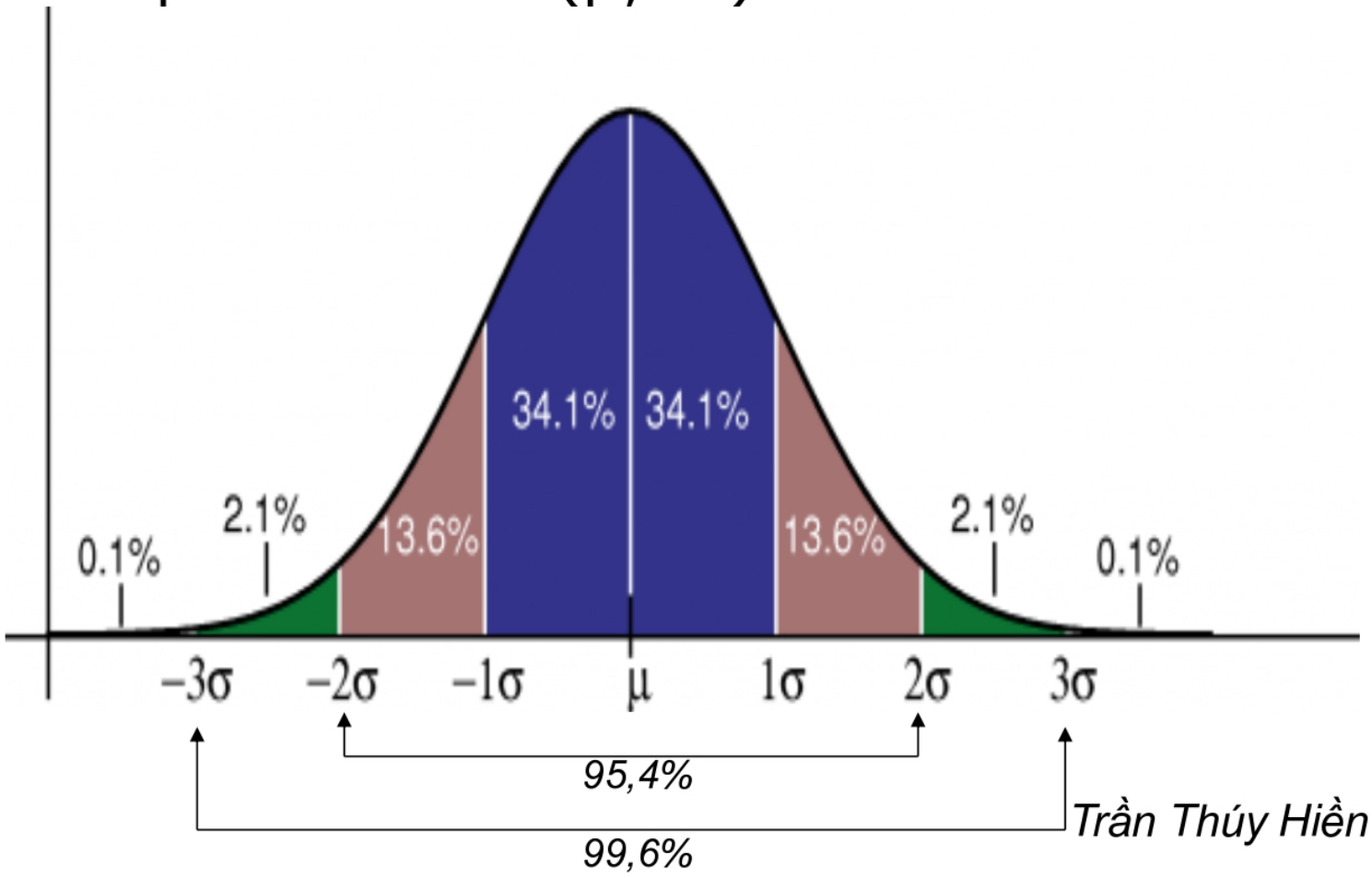
Từ đó, rút ra những suy diễn về quá trình hay về tập hợp các đơn vị được nghiên cứu.



4.1. Lý thuyết thống kê và Một số đại lượng thống kê

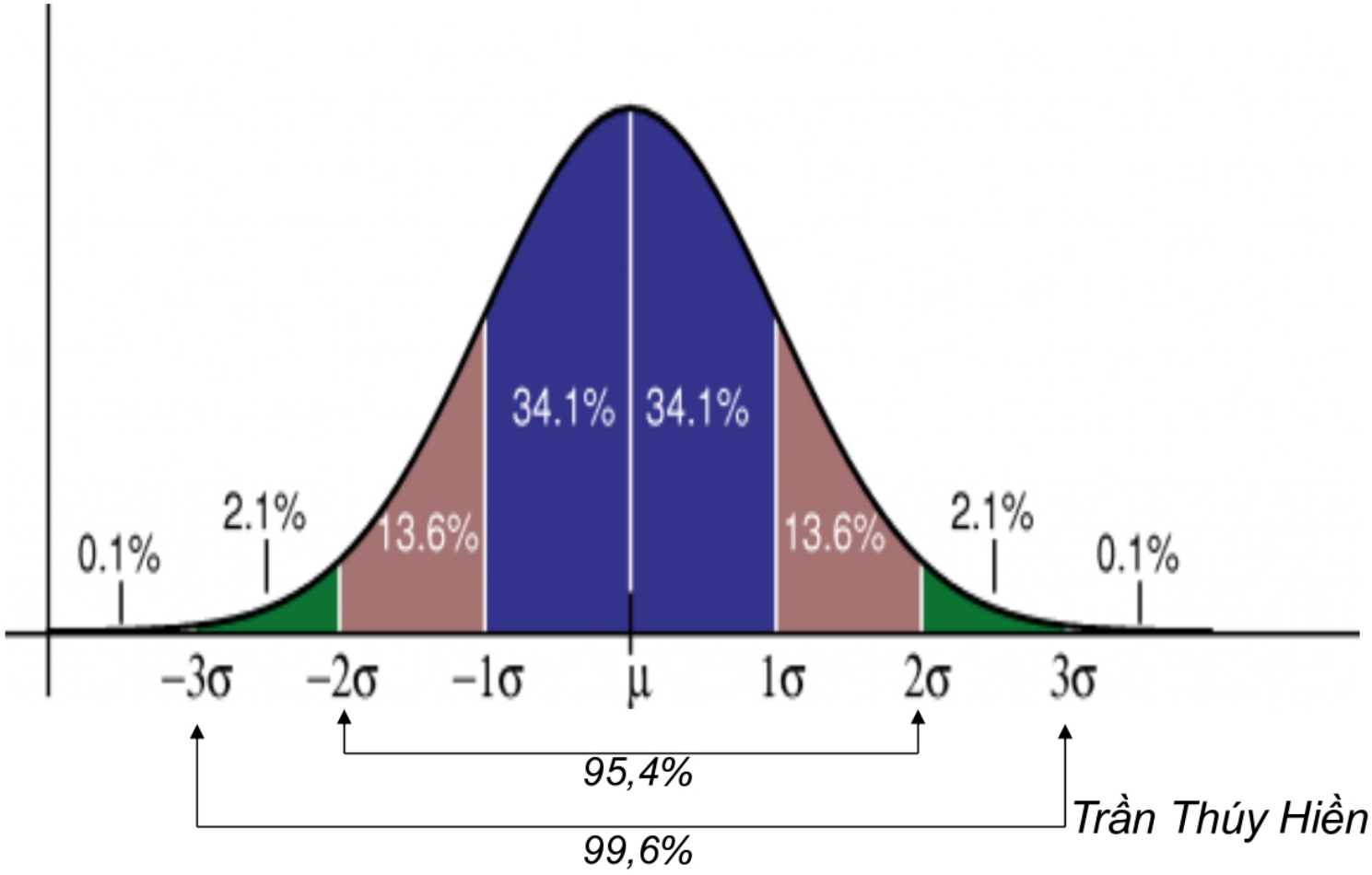
(4) Một số phân phối trong thống kê

a) Phân phối Chuẩn $N(\mu; \sigma^2)$



4.1. Lý thuyết thống kê và Một số đại lượng thống kê

Khoảng 95% các giá trị của biến thuộc khoảng (trung bình ±2. độ lệch chuẩn)



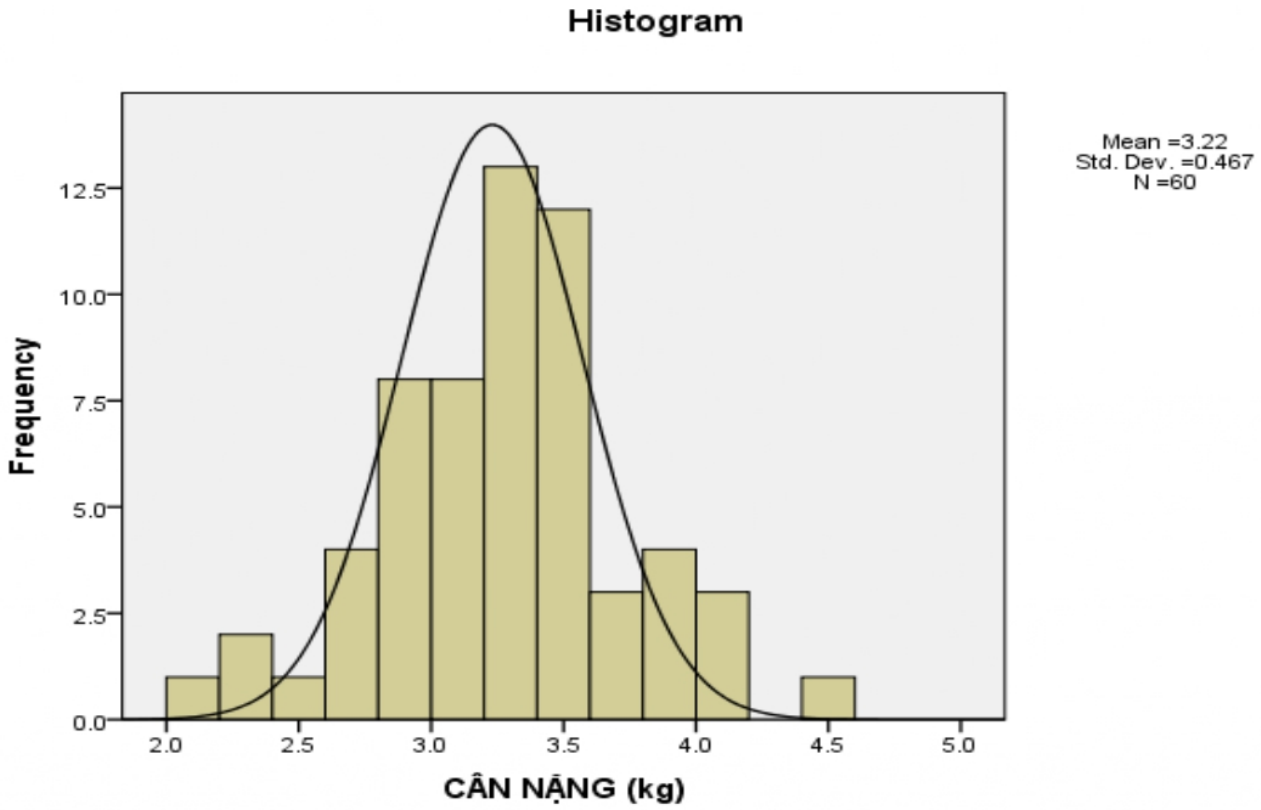
4.1. Lý thuyết thống kê và Một số đại lượng thống kê

Ví dụ. Cân nặng (kg) của 60 trẻ sơ sinh được chọn ngẫu nhiên trong số trẻ sinh ra tại một bệnh viện:

2,2	2,9	2,9	2,9	2,9	2,9	3,0	3,0	3,0	3,0
3,1	3,1	3,1	3,1	3,2	3,2	3,2	3,2	3,2	3,2
3,3	3,3	3,3	3,3	3,3	3,3	3,3	3,4	3,4	3,4
3,4	3,4	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,6
3,6	3,6	3,8	3,8	3,8	3,9	4,0	4,0	4,1	4,4

4.1. Lý thuyết thống kê và Một số đại lượng thống kê

Dữ liệu thỏa mãn tính chuẩn



Trần Thúy Hiền

4.1. Lý thuyết thống kê và Một số đại lượng thống kê

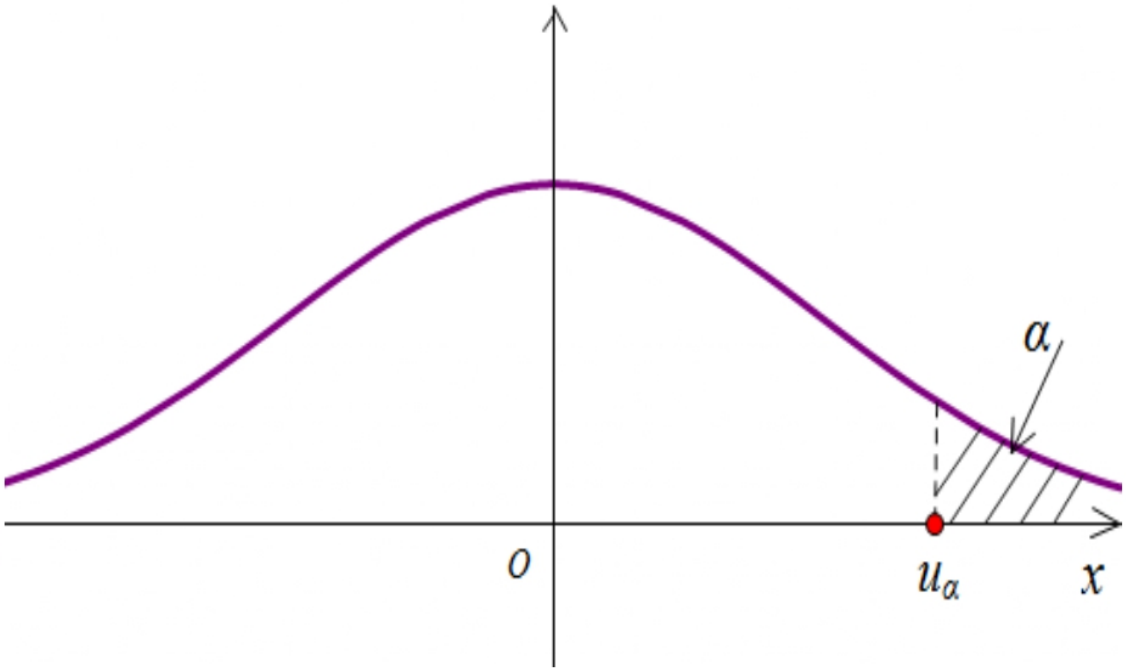
Phân phối Chuẩn hóa N(0; 1) và giá trị tới hạn u_α

$$\alpha = 0,05 \rightarrow u_\alpha = u_{0,05} = 1,65$$

$$\alpha = 0,025 \rightarrow u_\alpha = u_{0,025} = 1,96$$

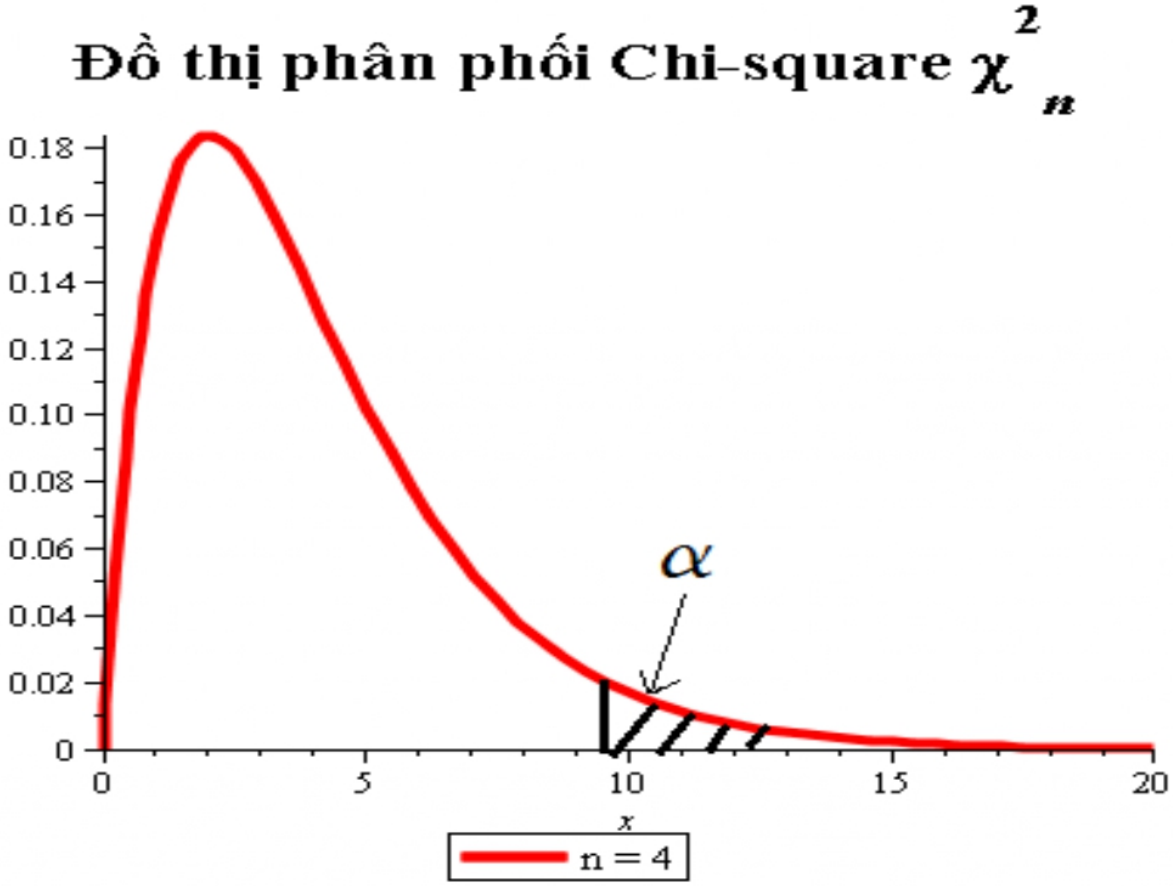
$$\alpha = 0,01 \rightarrow u_\alpha = u_{0,01} = 2,33$$

$$\alpha = 0,005 \rightarrow u_\alpha = u_{0,005} = 2,58$$



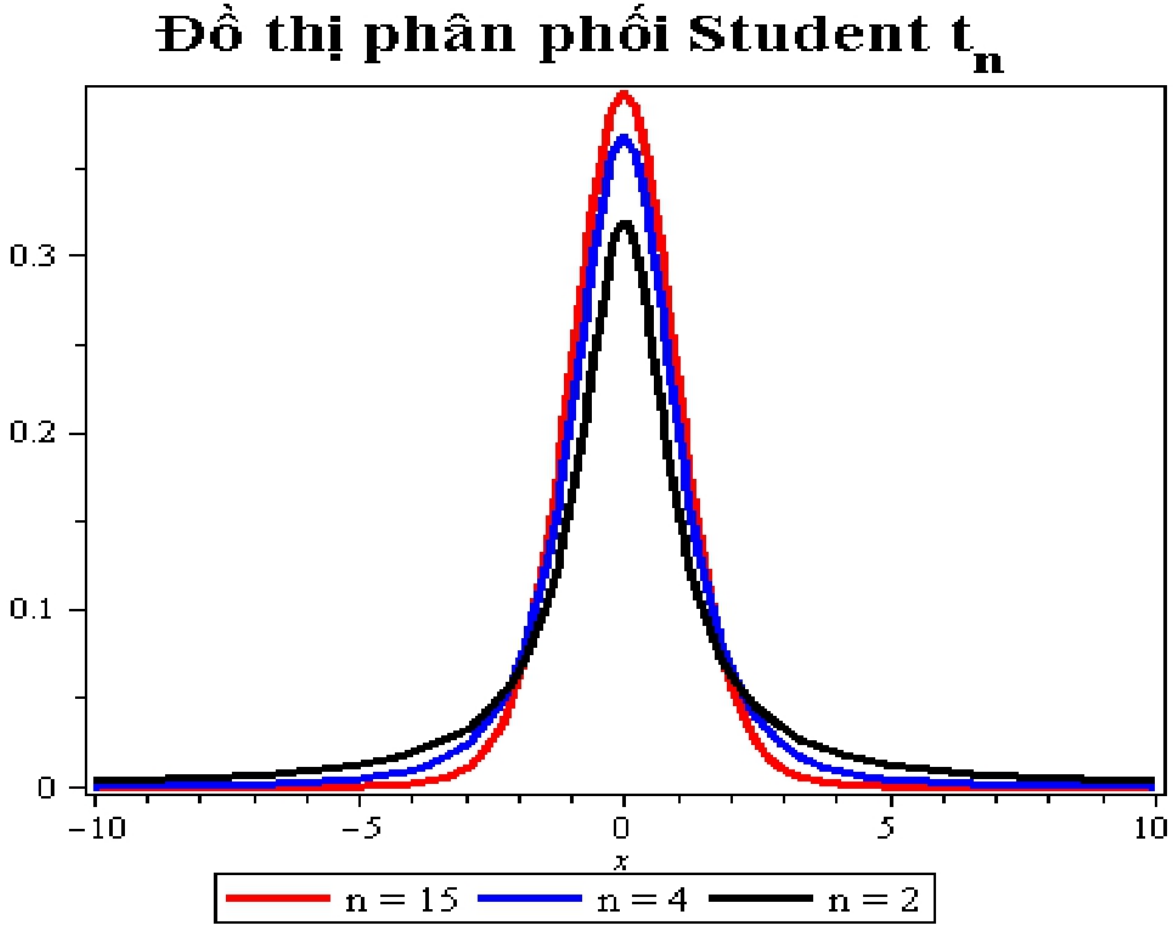
4.1. Lý thuyết thống kê và Một số đại lượng thống kê

b) Phân phối Chi – Square χ^2_n và giá trị tới hạn $\chi^2_n(\alpha)$

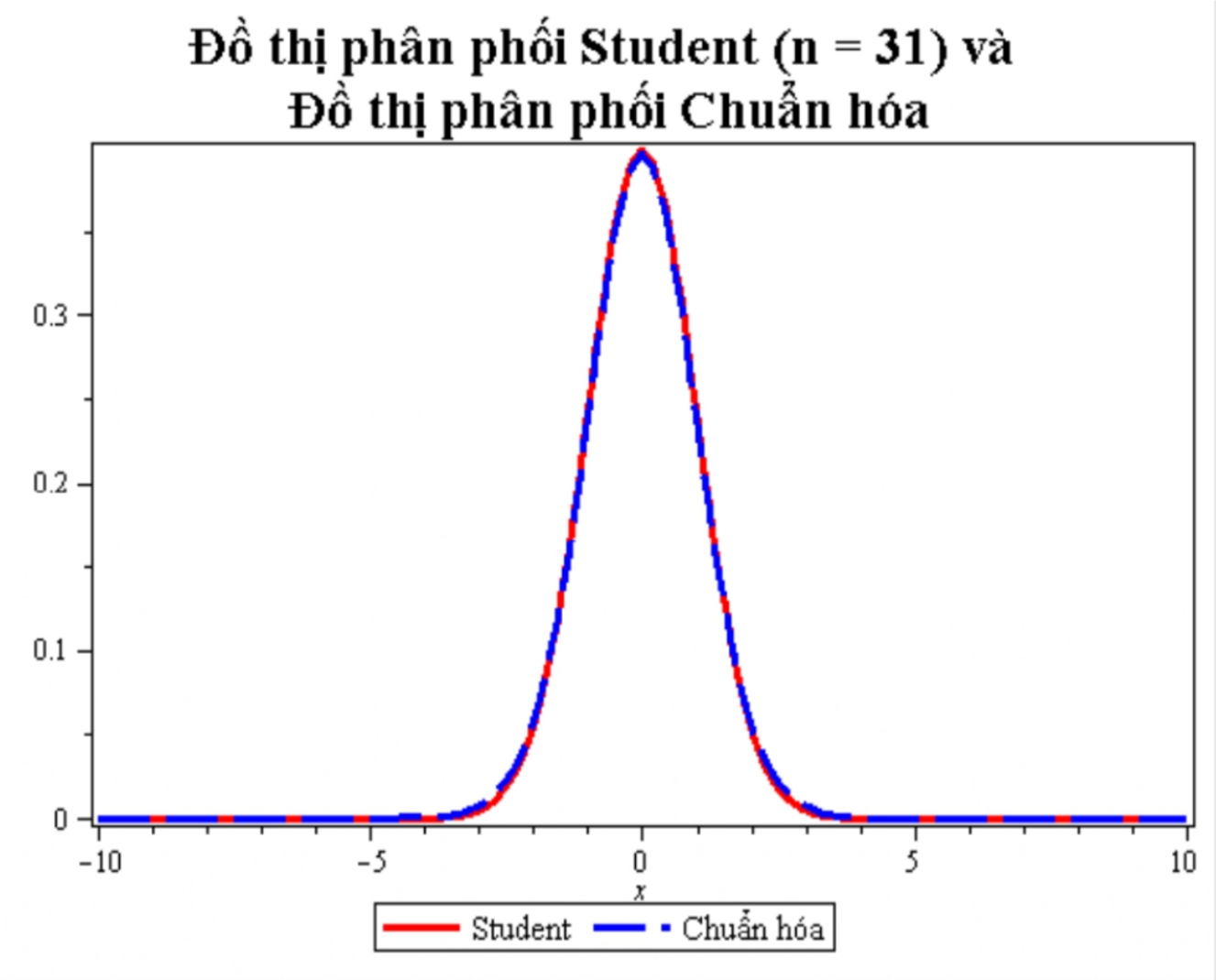


4.1. Lý thuyết thống kê và Một số đại lượng thống kê

c) Phân phối Student t_n và giá trị tới hạn $t_n(\alpha)$

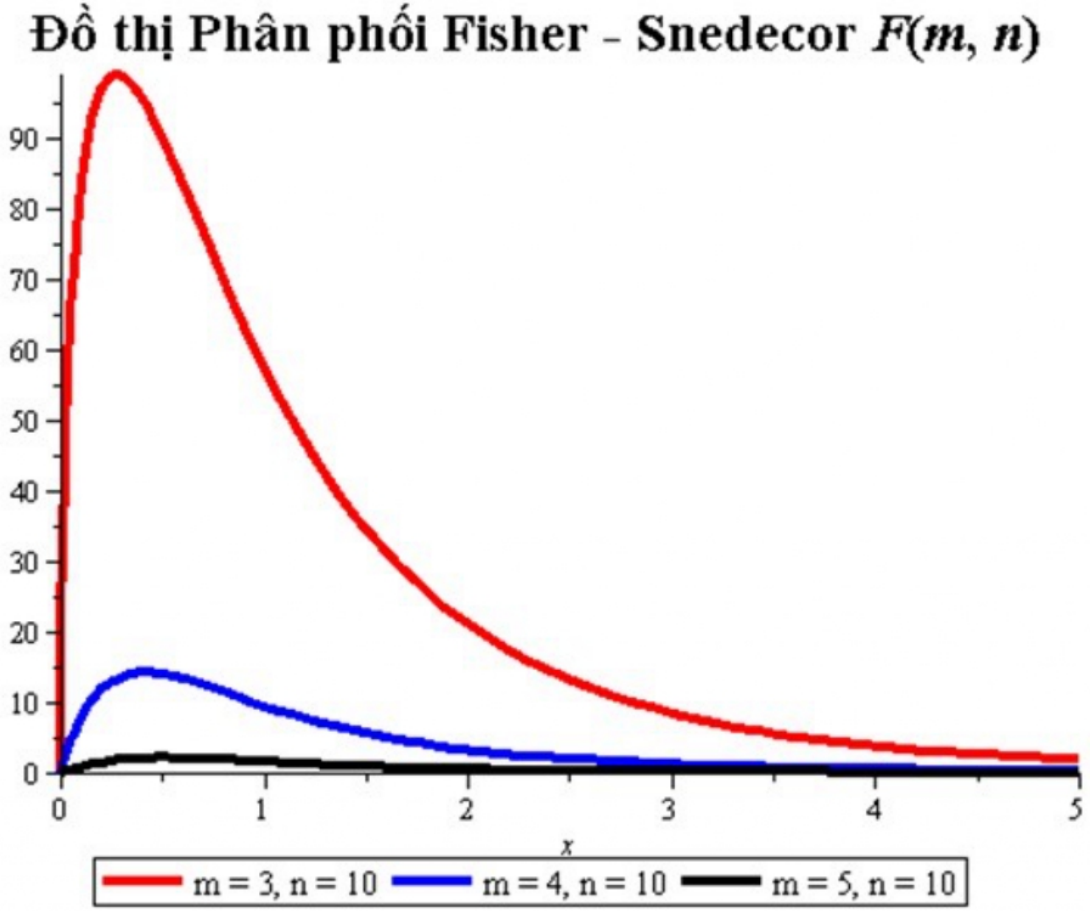


4.1. Lý thuyết thống kê và Một số đại lượng thống kê



4.1. Lý thuyết thống kê và Một số đại lượng thống kê

d) Phân phối Fisher – Snedecor F(m, n)



4.1. Lý thuyết thống kê và Một số đại lượng thống kê

(5) Tóm tắt, mô tả dữ liệu bằng Bảng, Biểu đồ

Khoang Chieu cao

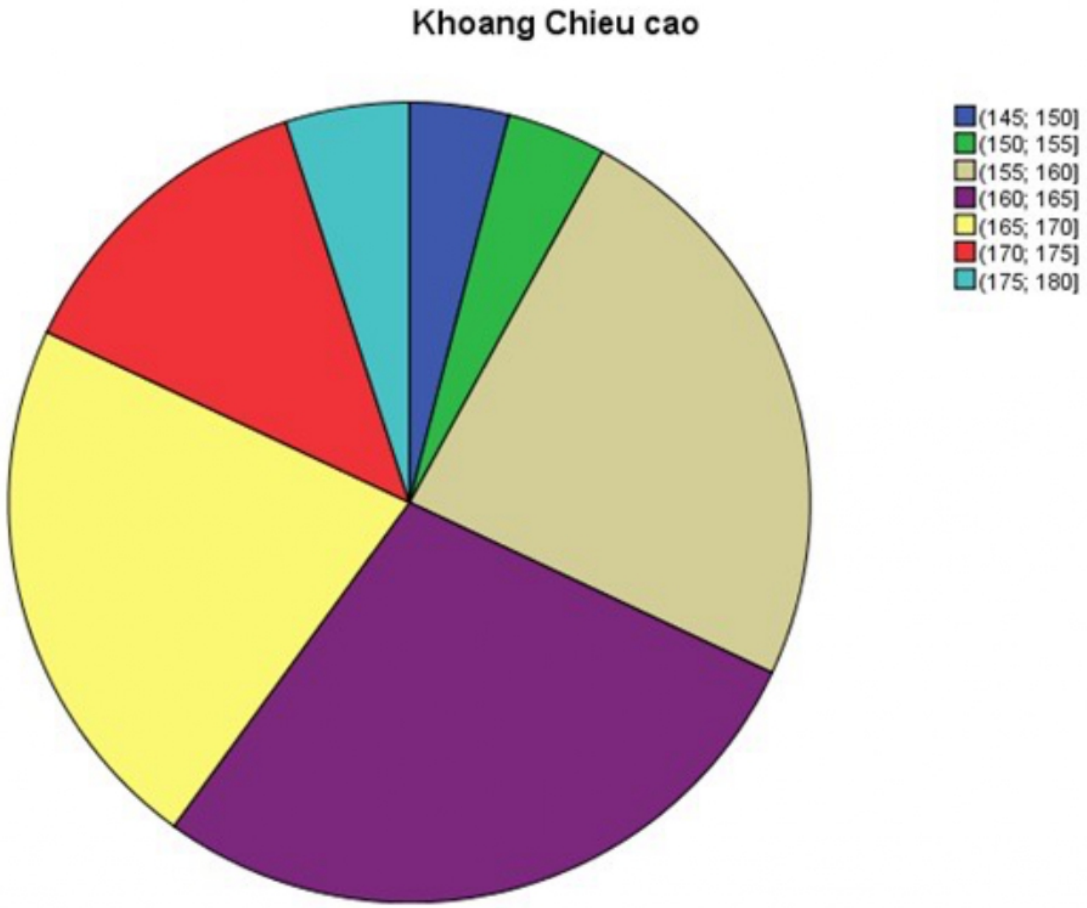
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	(145; 150]	4	4.0	4.0	4.0
	(150; 155]	4	4.0	4.0	8.0
	(155; 160]	24	24.0	24.0	32.0
	(160; 165]	28	28.0	28.0	60.0
	(165; 170]	22	22.0	22.0	82.0
	(170; 175]	13	13.0	13.0	95.0
	(175; 180]	5	5.0	5.0	100.0
	Total	100	100.0	100.0	

4.1. Lý thuyết thống kê và Một số đại lượng thống kê

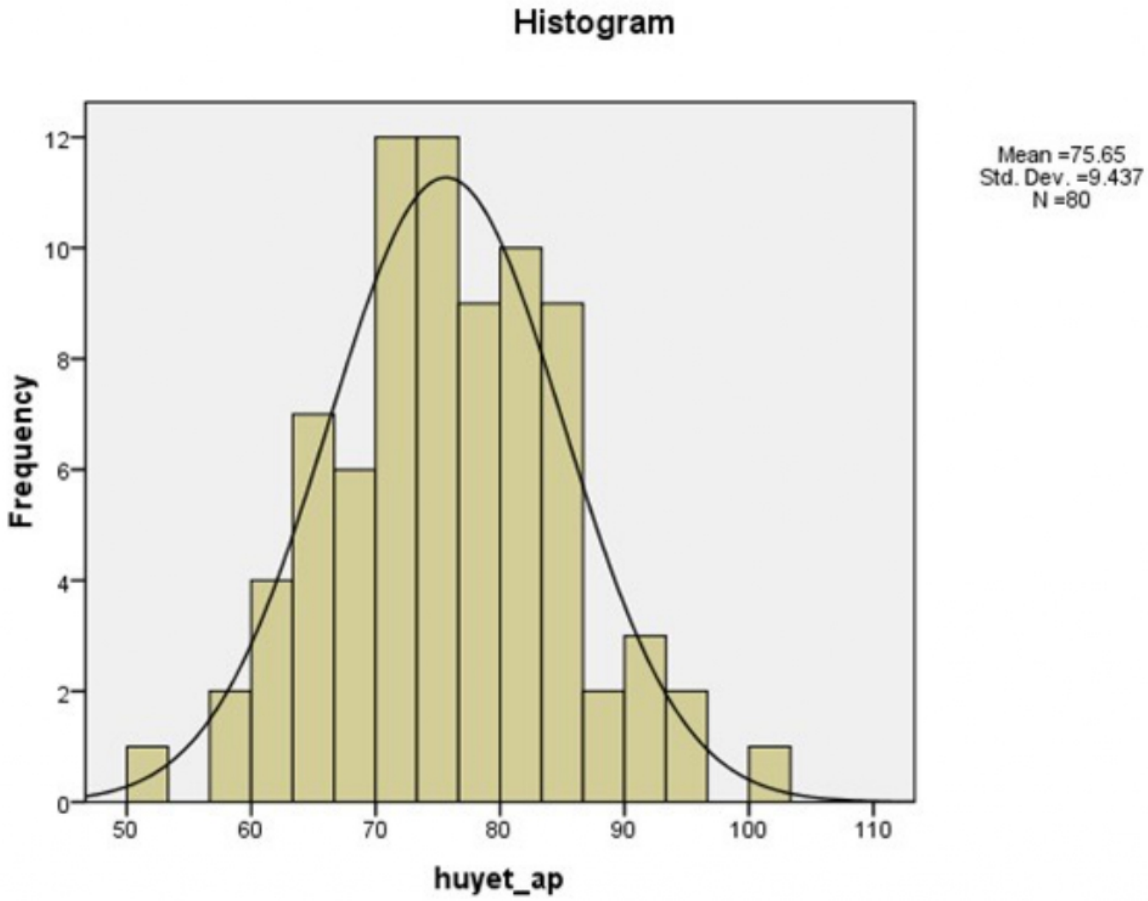
suame * nhiemHP Crosstabulation

			nhiemHP		Total
			0	1	
suame	0	Count	5	10	15
		Expected Count	9.0	6.0	15.0
		% within suame	33.3%	66.7%	100.0%
	1	Count	13	2	15
		Expected Count	9.0	6.0	15.0
		% within suame	86.7%	13.3%	100.0%
Total		Count	18	12	30
		Expected Count	18.0	12.0	30.0
		% within suame	60.0%	40.0%	100.0%

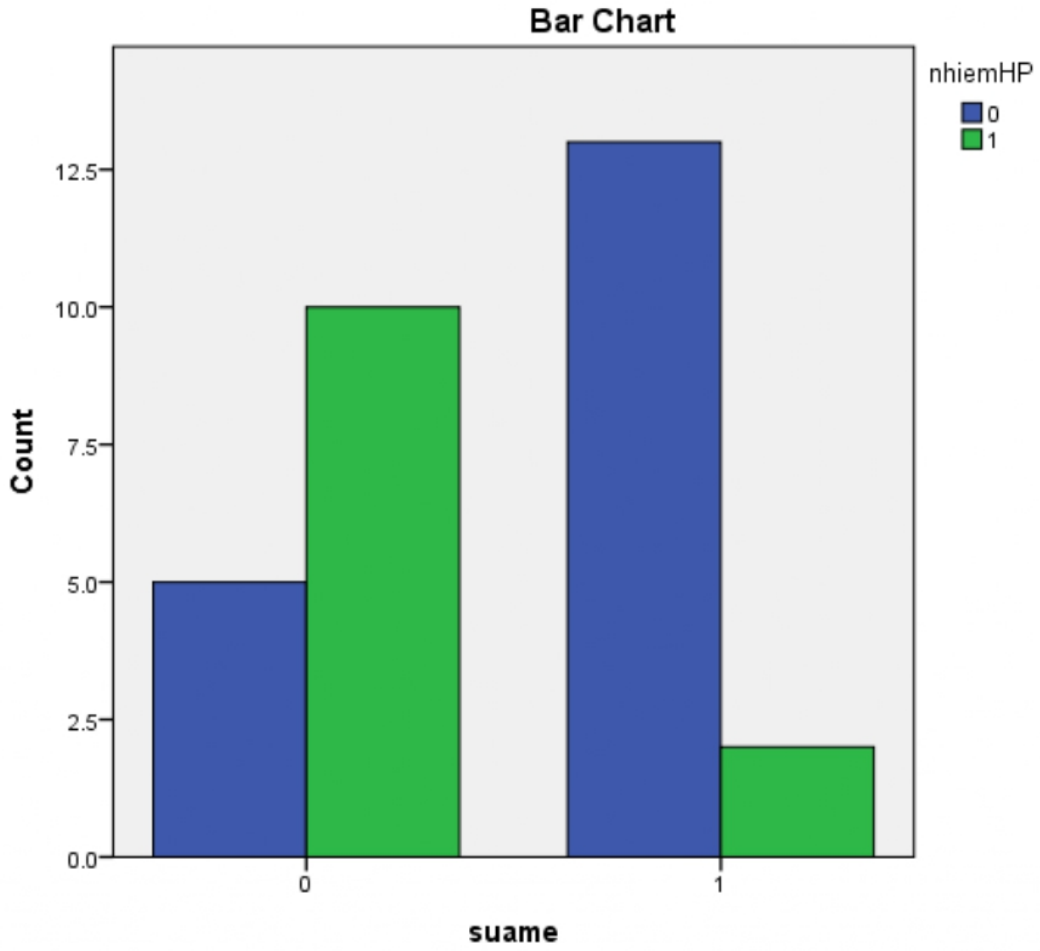
4.1. Lý thuyết thống kê và Một số đại lượng thống kê



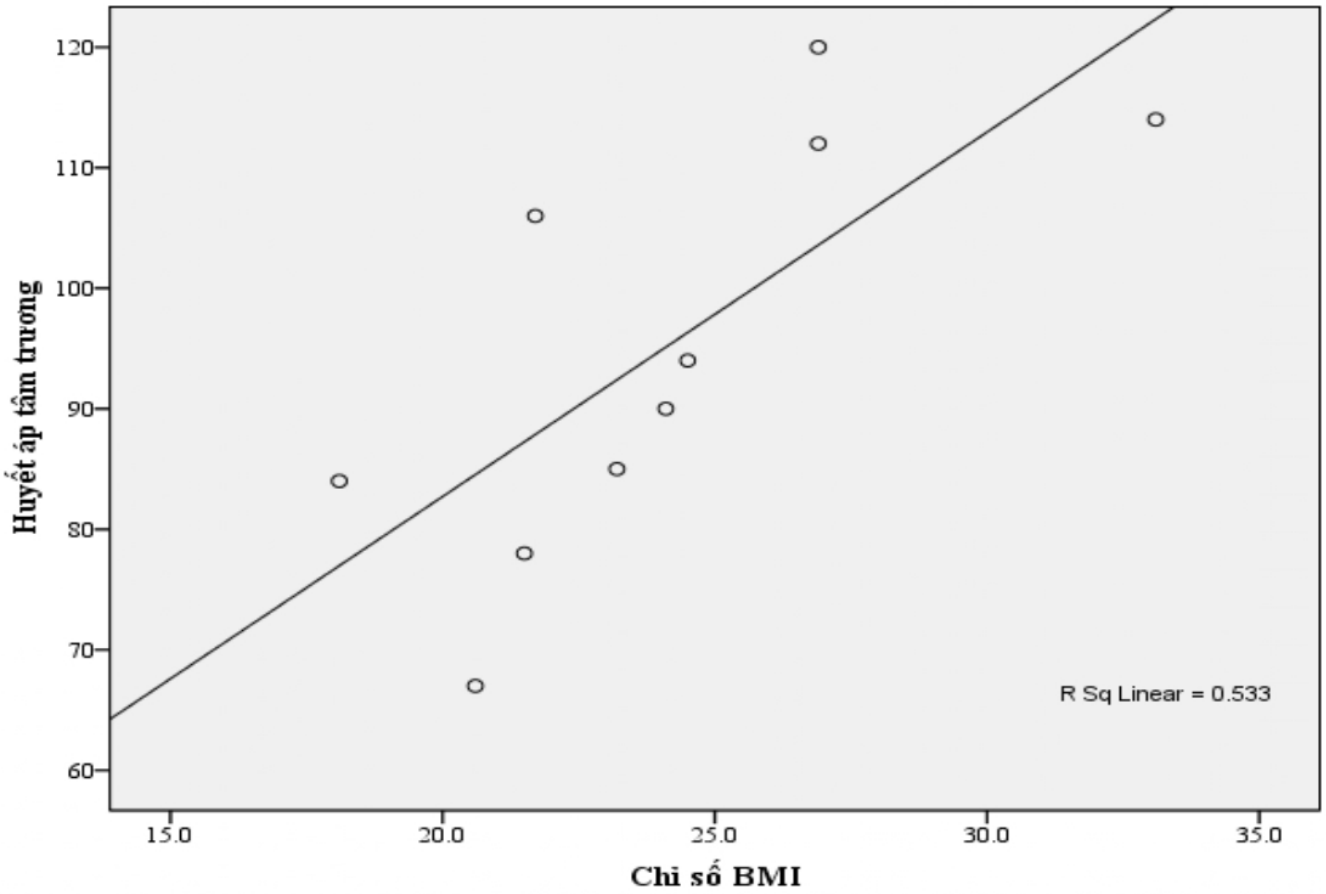
4.1. Lý thuyết thống kê và Một số đại lượng thống kê



4.1. Lý thuyết thống kê và Một số đại lượng thống kê



4.1. Lý thuyết thống kê và Một số đại lượng thống kê



4.1. Lý thuyết thống kê và Một số đại lượng thống kê

4.1.2. Một số đại lượng thống kê

Tiêu thức	Tham số Quần thể	Thống kê Mẫu
1. Trung bình	μ	$\bar{X}$
2. Phương sai	σ^2	S^2
3. Độ lệch chuẩn	σ	S
4. Tỷ lệ	p	f

4.1. Lý thuyết thống kê và Một số đại lượng thống kê

1. Đại lượng đo lường độ tập trung

- Giá trị trung bình (Mean)
- Trung vị (Median)
- Số trội (Mode)

2. Đại lượng đo lường độ phân tán

- Khoảng biến thiên (Range)
- Phương sai (Variance)
- Độ lệch chuẩn (Standard Deviation - Std.Deviation)
- Sai số chuẩn (Standard Error of Mean – SEM)
- Khoảng tin cậy (Confident interval)

3. Đại lượng đo lường dạng hình phân phối

- Hệ số đối xứng (Skewness):
- Hệ số nhọn (Kurtosis):

4.1. Lý thuyết thống kê và Một số đại lượng thống kê

Statistics

huyet_ap

N	Valid	80
	Missing	0
Mean		75.65
Median		76.00
Mode		70

4.1. Lý thuyết thống kê và Một số đại lượng thống kê

- Giá trị trung bình (Mean): $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$
  - Chịu ảnh hưởng của các giá trị quan sát.
  - Thường sử dụng để so sánh 2 hay nhiều tổng thể.
- Số trội (Mode):
  - Là giá trị quan sát có tần suất xuất hiện lớn nhất trong dãy số quan sát.
  - Mod không bị ảnh hưởng bởi các giá trị đột biến.

4.1. Lý thuyết thống kê và Một số đại lượng thống kê

- Trung vị (Median):
  - Dãy số quan sát được sắp thứ tự từ nhỏ đến lớn thì Med là điểm giữa, chia dãy số thành 2 phần bằng nhau.
  - Med không bị ảnh hưởng bởi các giá trị đột biến.

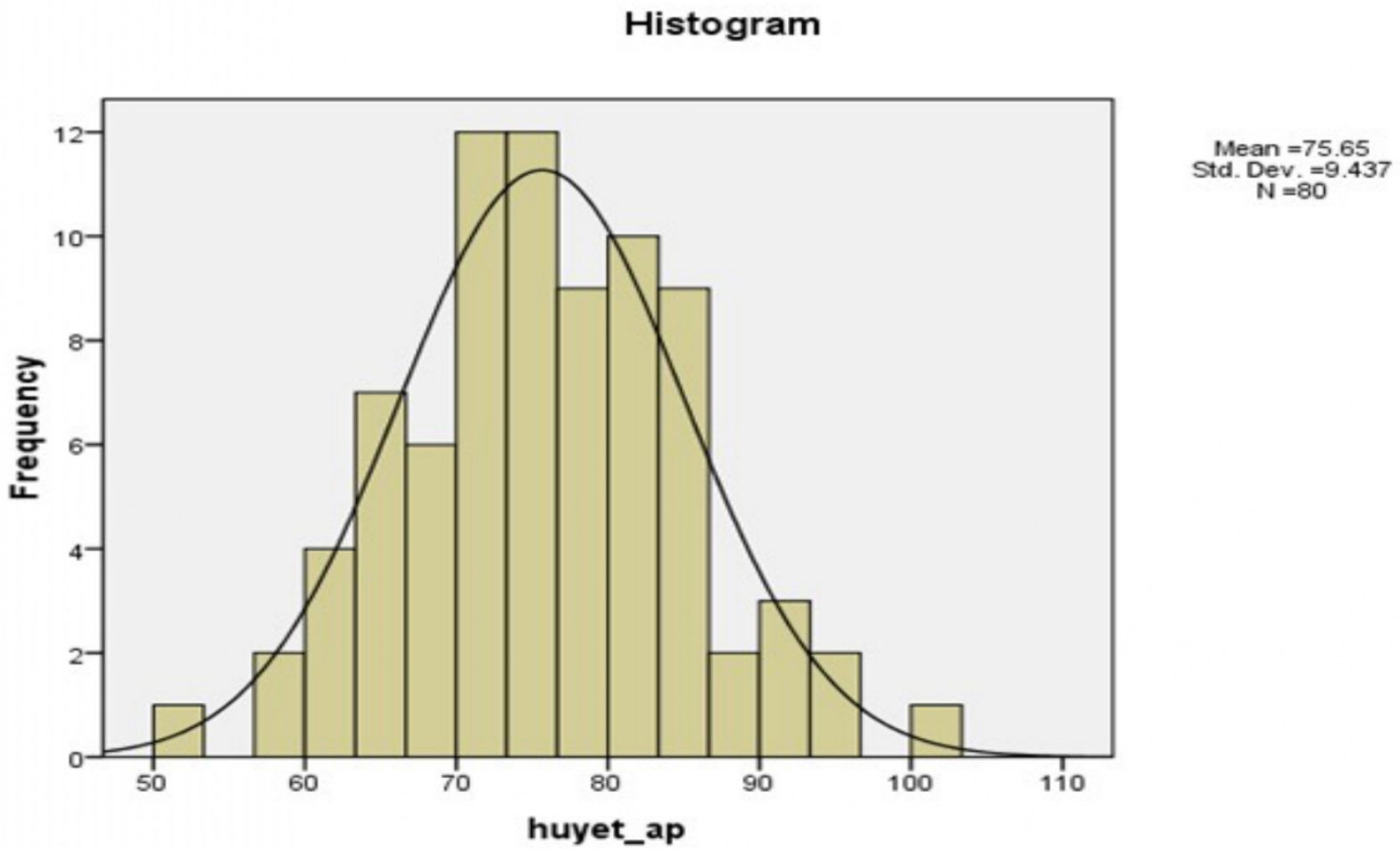
Ví dụ. Số liệu về chỉ số đau (pain index) của 11 bệnh nhân thấp khớp như sau

0,05 0,15 0,35 0,25 0,20 0,05 0,10
0,05 0,30 0,05 0,25

0.05,	0.05,	0.05,	0.05,	0.10,	0.15,	0.20,	0.25,	0.25,	0.30,	0.35
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)

4.1. Lý thuyết thống kê và Một số đại lượng thống kê

Dữ liệu thỏa mãn tính chuẩn: giá trị Mean, Median, Mode



4.1. Lý thuyết thống kê và Một số đại lượng thống kê

Statistics

huyet_ap

N	Valid	80
	Missing	0
Std. Error of Mean		1.055
Std. Deviation		9.437
Variance		89.066
Range		47
Minimum		53
Maximum		100
Percentiles	25	69.25
	50	76.00
	75	81.00

4.1. Lý thuyết thống kê và Một số đại lượng thống kê

- Khoảng biến thiên (Range): $R = X_{\max} - X_{\min}$
- Phương sai (Variance): $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{X})^2$
- Độ lệch chuẩn (Standard deviation - SD):

$$SD = S = \sqrt{S^2}$$

- Sai số chuẩn (Standard Error of Mean – Se):

$$Se = \frac{S}{\sqrt{n}}$$

4.1. Lý thuyết thống kê và Một số đại lượng thống kê

- Hệ số bất đối xứng (Skewness):
  - Kiểm tra tính chuẩn: nếu nằm ngoài đoạn [-2; 2] thì không xem là chuẩn.
- Hệ số nhọn (Kurtosis):
  - So sánh đường cong quan sát với dạng đường cong phân phối chuẩn.
  - Kiểm tra tính chuẩn: nếu nằm ngoài đoạn [-2; 2] thì không xem là chuẩn.

4.1. Lý thuyết thống kê và Một số đại lượng thống kê

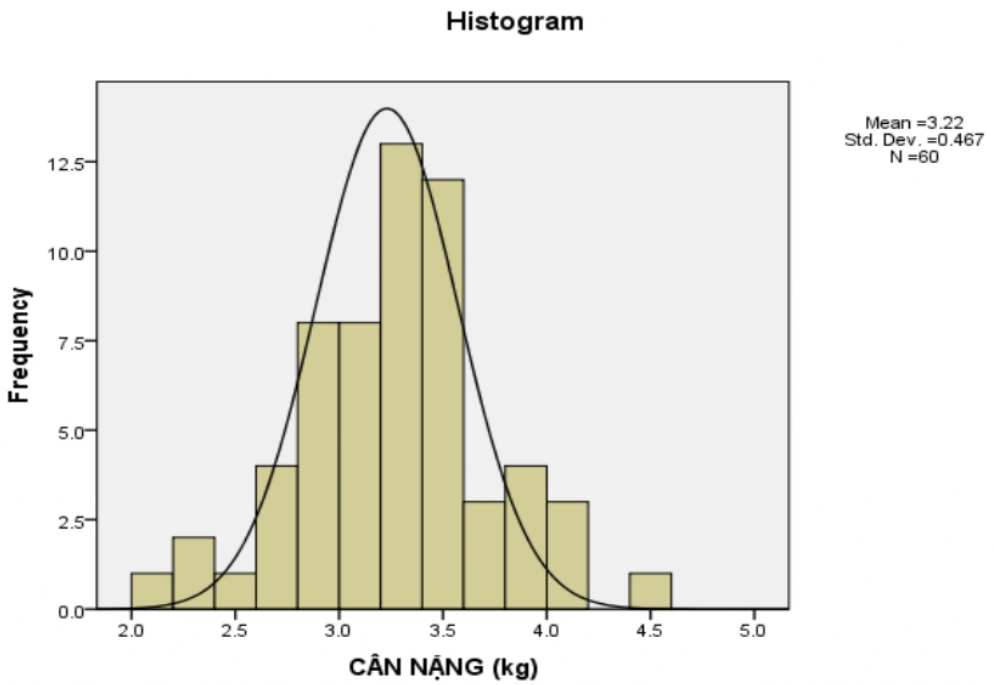
Frequencies

Statistics

CÂN NẶNG (kg)

N	Valid	60
	Missing	0
Mean		3.217
Std. Error of Mean		.0603
Median		3.250
Mode		3.3[a]
Std. Deviation		.4673
Variance		.218
Skewness		-.140
Std. Error of Skewness		.309
Kurtosis		.331
Std. Error of Kurtosis		.608
Range		2.3
Minimum		2.1
Maximum		4.4
Sum		193.0

a. Multiple modes exist. The smallest value is shown



Trần Thúy Hiền

4.1. Lý thuyết thống kê và Một số đại lượng thống kê

Phân phối chuẩn:

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
CÂN NẶNG (kg)	.089	60	.200*	.983	60	.588

a.Lilliefors Significance Correction

*.This is a lower bound of the true significance.

4.2. Bảng tần số (Frequencies)

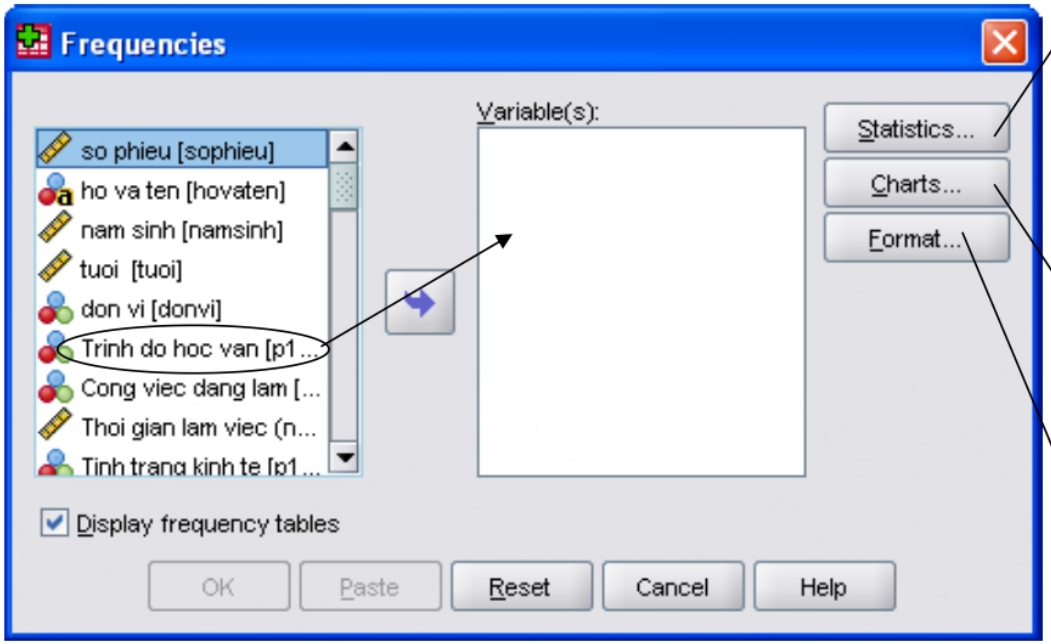
1. Chức năng

- Khảo sát dữ liệu thông qua các công cụ: tần suất xuất hiện, phần trăm, phần trăm tích lũy.
- Cung cấp các phép đo lường thống kê: độ tập trung, độ phân tán, tứ phân vị, các bách phân vị, phân phối dữ liệu.
- Giúp phát hiện những sai sót trong dữ liệu: giá trị bất thường, giá trị mã hóa bất thường.

4.2. Bảng tần số (Frequencies)

2. Thao tác thực hiện

- Analyze/Descriptive Statistics/Frequencies



Truy xuất các công cụ đo lường giá trị thống kê của dữ liệu

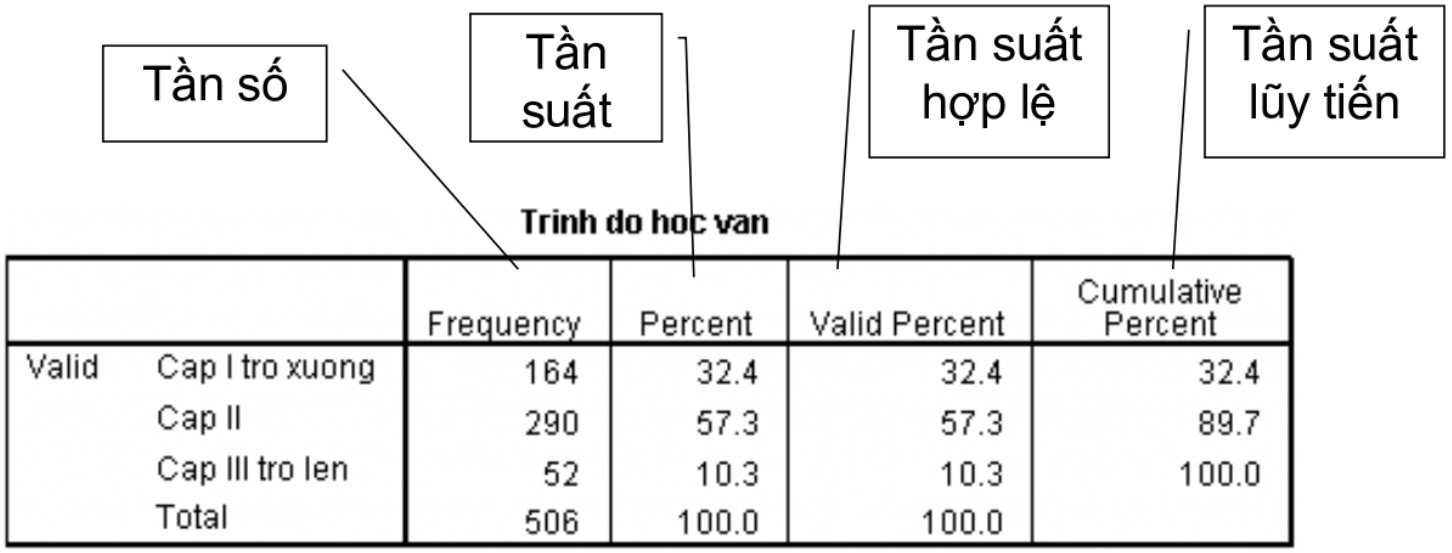
Vẽ đồ thị cho dữ liệu

Định ra kiểu hiển thị của dữ liệu

4.2. Bảng tần số (Frequencies)

2. Thao tác thực hiện

- Click nút OK sẽ cho kết quả



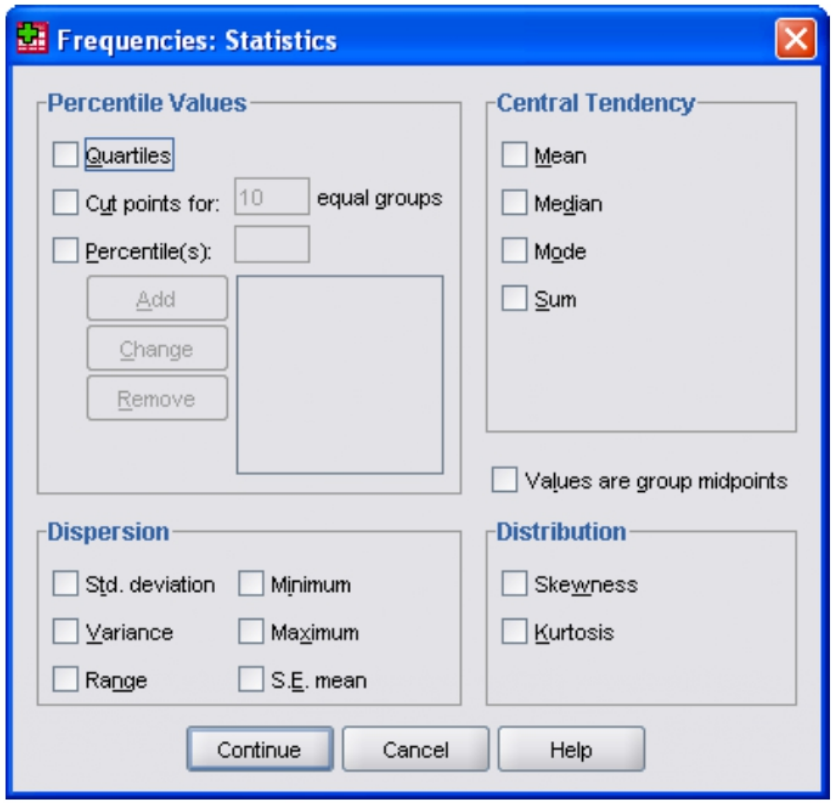
Trình độ học vấn

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Cap I tro xuong	164	32.4	32.4	32.4
	Cap II	290	57.3	57.3	89.7
	Cap III tro len	52	10.3	10.3	100.0
	Total	506	100.0	100.0	

4.2. Bảng tần số (Frequencies)

2. Thao tác thực hiện

- Click nút Statistics… xuất hiện hộp thoại chứa các công cụ đo lường giá trị thống kê của dữ liệu



4.3. Bảng đại lượng thống kê mô tả (Descriptives)

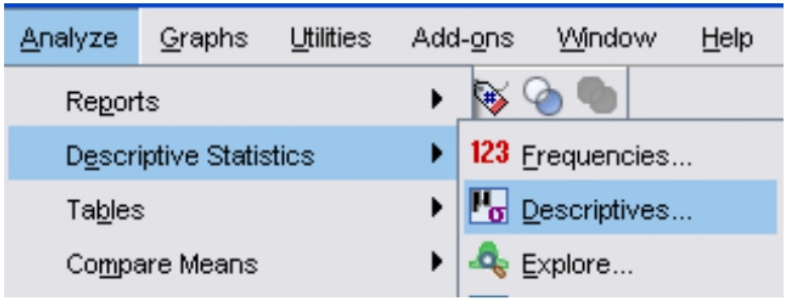
1. Chức năng

- Tóm tắt dữ liệu (dữ liệu định lượng)
- Thể hiện xu hướng tập trung của dữ liệu (thông qua giá trị trung bình của các giá trị trong biến – Mean)
- Mô tả sự phân tán của dữ liệu (thông qua phương sai và độ lệch chuẩn)

4.3. Bảng đại lượng thống kê mô tả (Descriptives)

2. Thao tác thực hiện

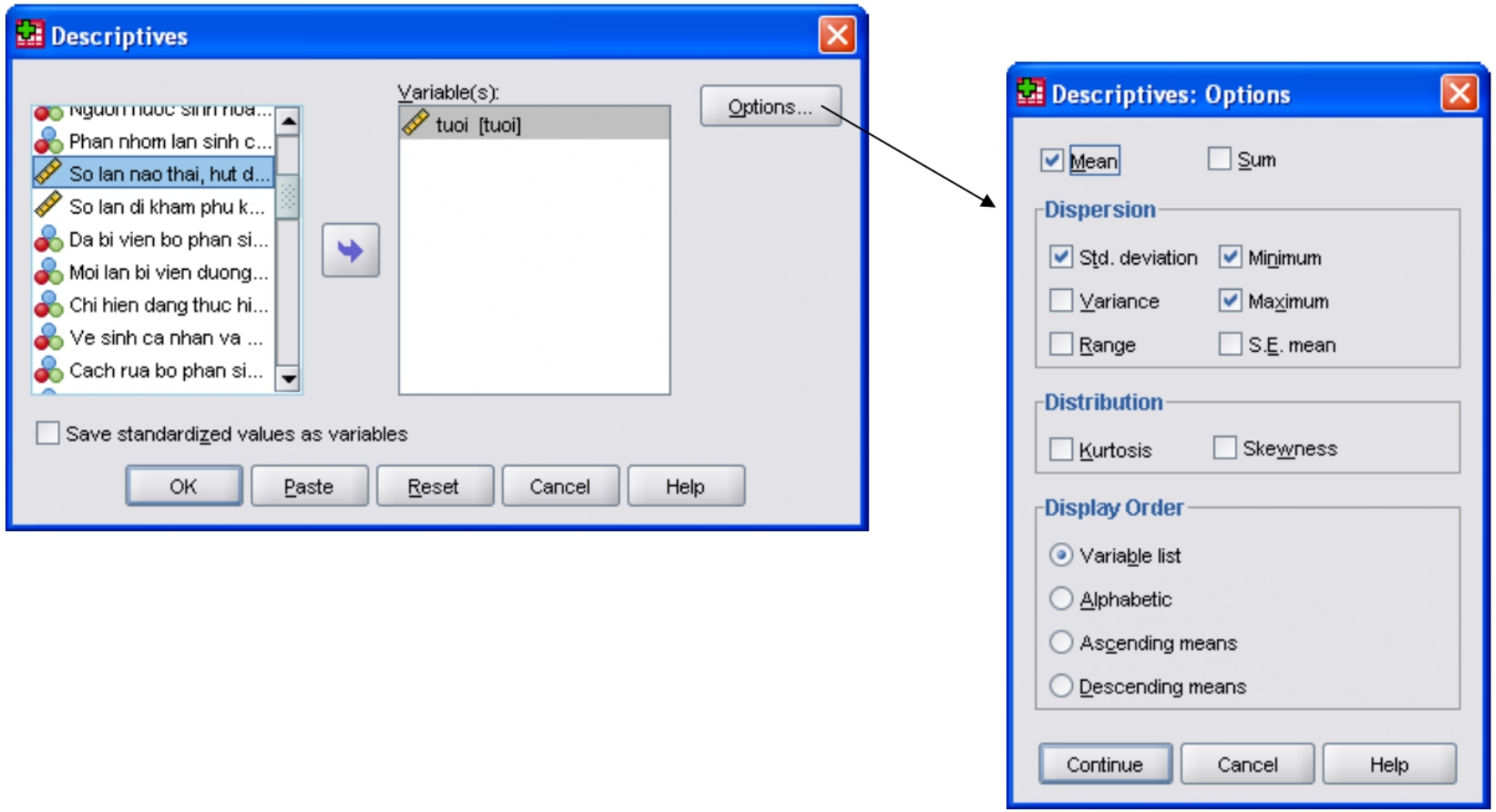
- Analyze/Descriptives Statistics/ Descriptives…



- Chuyển các biến cần tóm tắt vào hộp thoại Variables
- Click nút Options để chọn các đại lượng thống kê cần mô tả: giá trị trung bình, giá trị nhỏ nhất, giá trị lớn nhất, phương sai và độ lệch chuẩn, …

4.3. Bảng đại lượng thống kê mô tả (Descriptives)

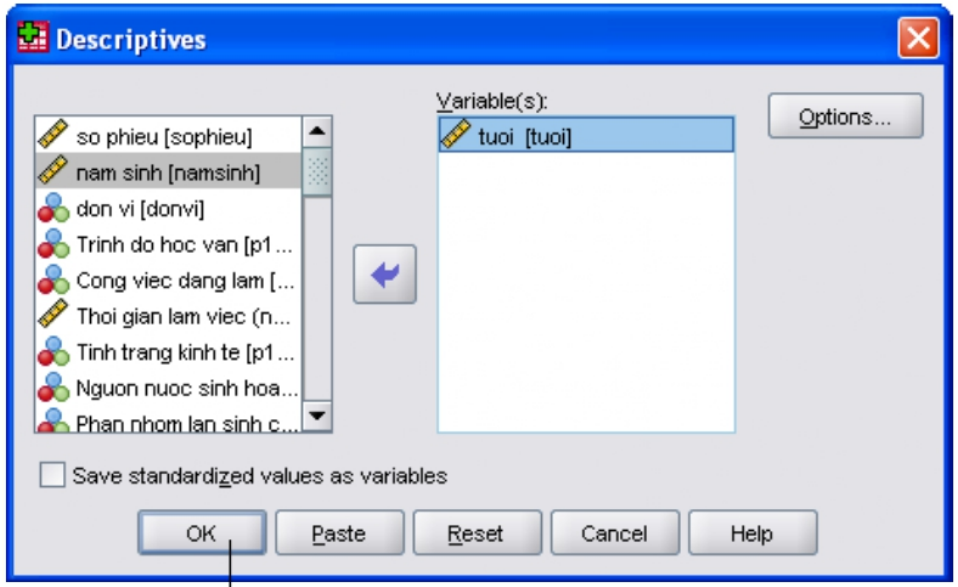
2. Thao tác thực hiện



Trần Thúy Hiền

4.3. Bảng đại lượng thống kê mô tả (Descriptives)

2. Thao tác thực hiện



Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
tuoi	506	25	57	42.57	7.025
Valid N (listwise)	506				

Trần Thúy Hiền

4.4. Khám phá dữ liệu (Explore)

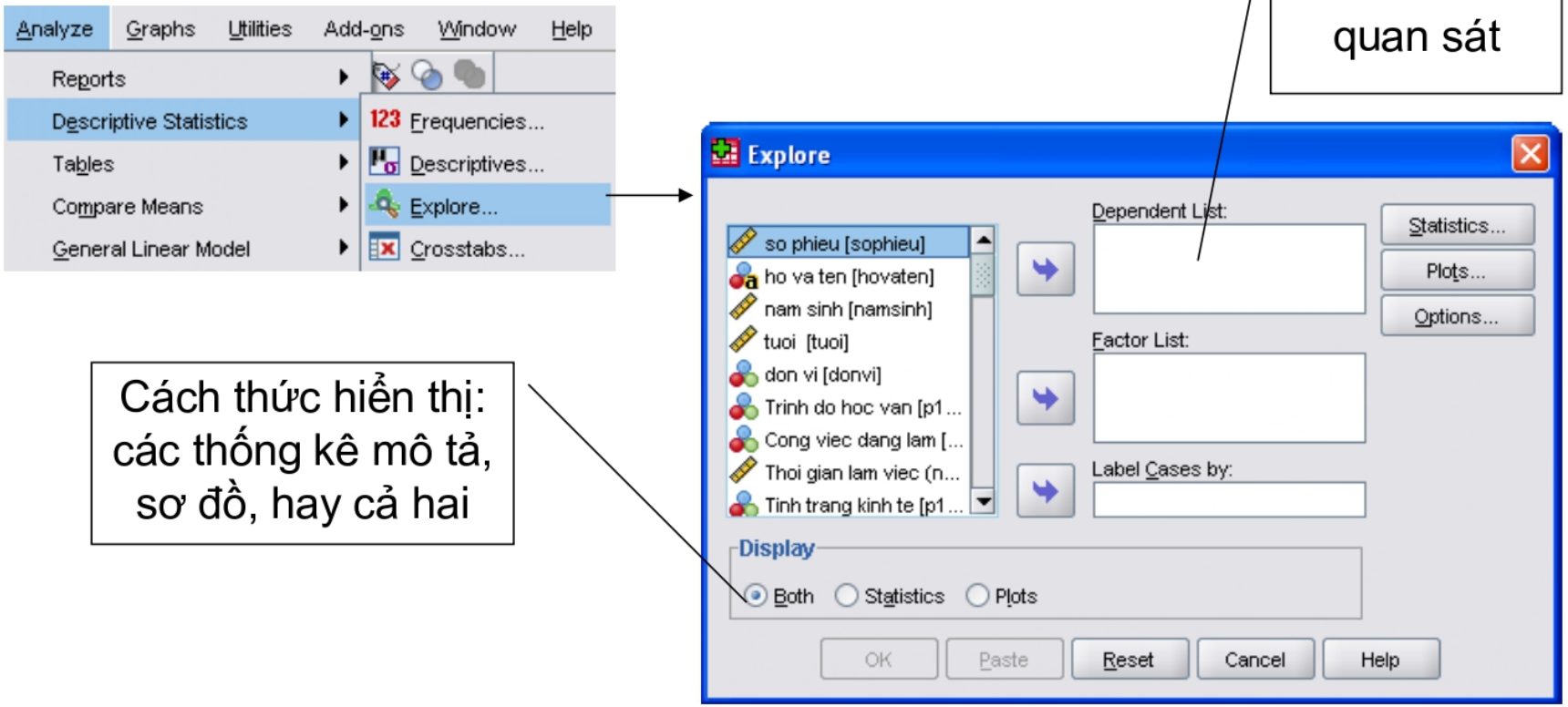
1. Chức năng

- Công cụ Explore cung cấp khả năng xem xét và kiểm tra dữ liệu:
  - Phát hiện các sai sót, các giá trị khác biệt
  - Nhận dạng dữ liệu để tìm phương pháp phân tích thích hợp
- Tính toán các đại lượng thống kê mô tả cho mọi trường hợp trong dữ liệu hoặc cho các nhóm con của chúng
- Tạo biểu đồ, hình dáng biểu đồ giúp nhận dạng về dạng phân phối

4.4. Khám phá dữ liệu (Explore)

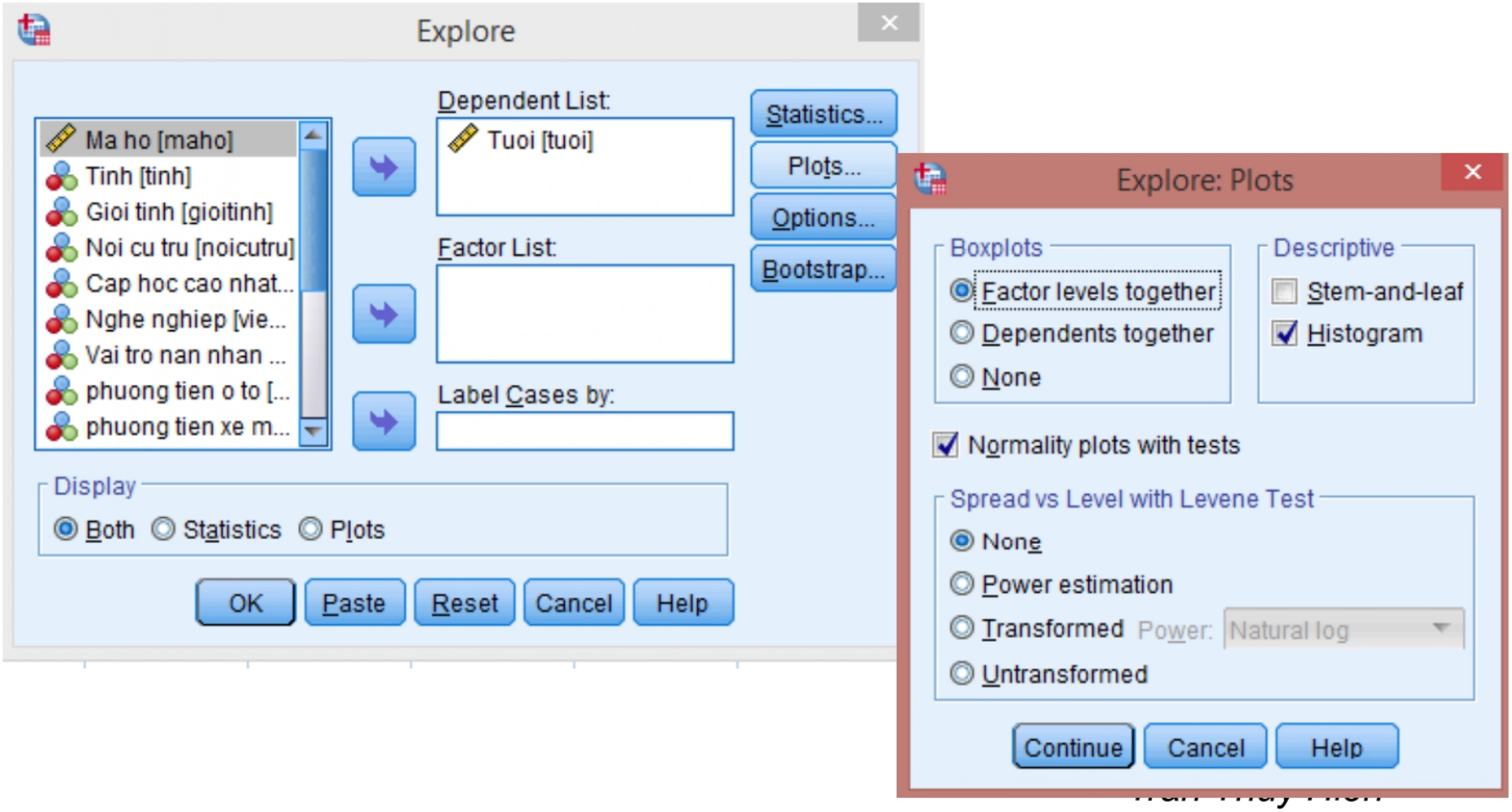
2. Thao tác thực hiện

- Analyze/Descriptive Statistics/Explore…



4.4. Khám phá dữ liệu (Explore)

2. Thao tác thực hiện



4.4. Khám phá dữ liệu (Explore)

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Tuoi	50	100.0%	0	0.0%	50	100.0%

Descriptives

			Statistic	Std. Error
Tuoi	Mean		36.36	2.309
	95% Confidence Interval for Mean	Lower Bound	31.72	
		Upper Bound	41.00	
	5% Trimmed Mean		35.99	
	Median		32.00	
	Variance		266.562	
	Std. Deviation		16.327	
	Minimum		11	
	Maximum		67	
	Range		56	
	Interquartile Range		28	
	Skewness		.269	.337
	Kurtosis		-1.216	.662

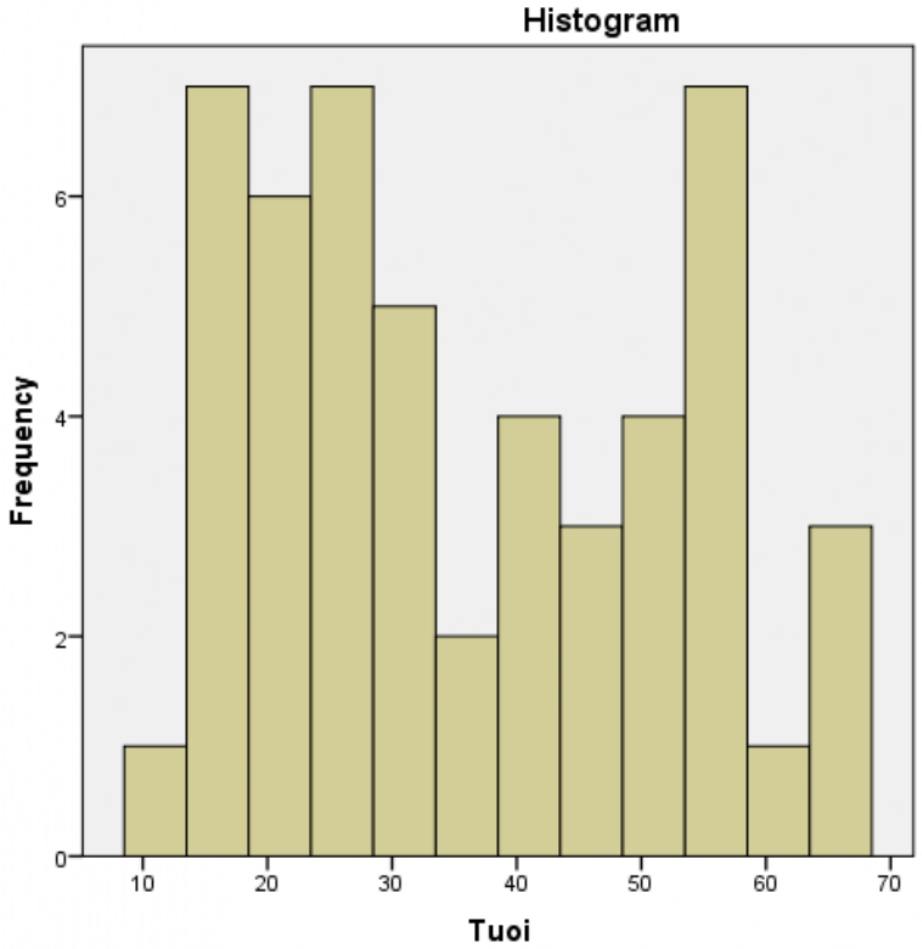
4.4. Khám phá dữ liệu (Explore)

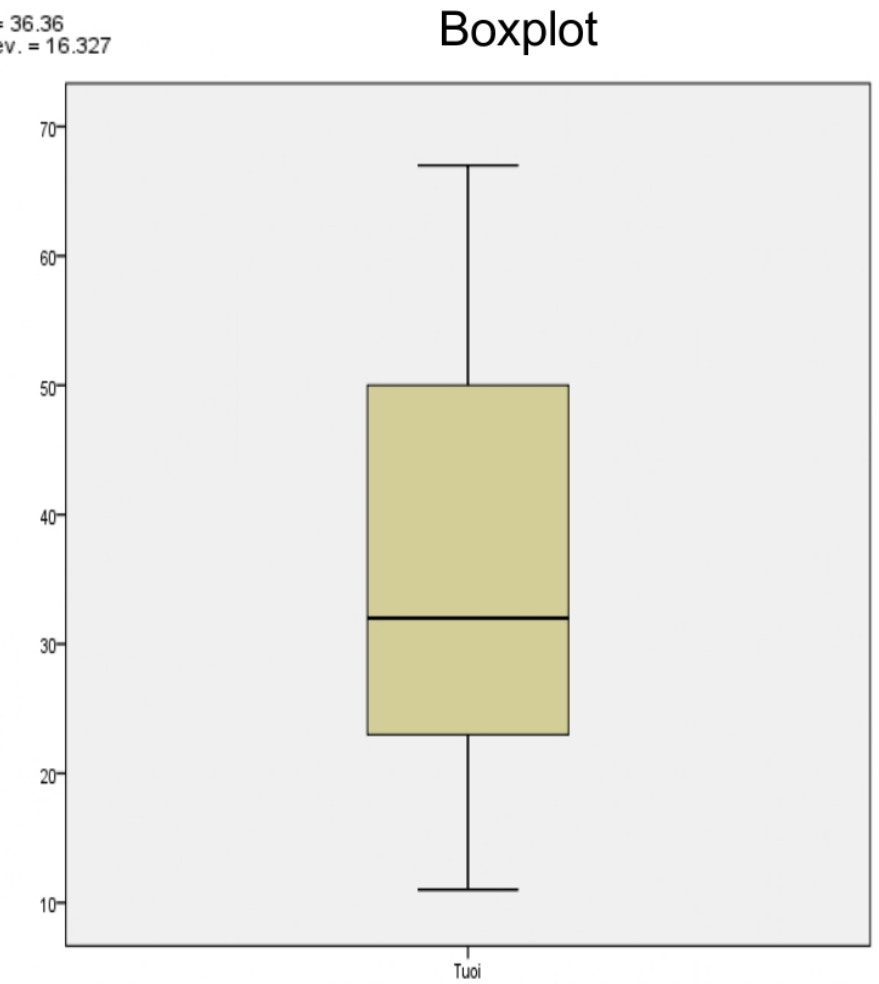
Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Tuoi	.134	50	.025	.938	50	.011

a. Lilliefors Significance Correction

4.4. Khám phá dữ liệu (Explore)

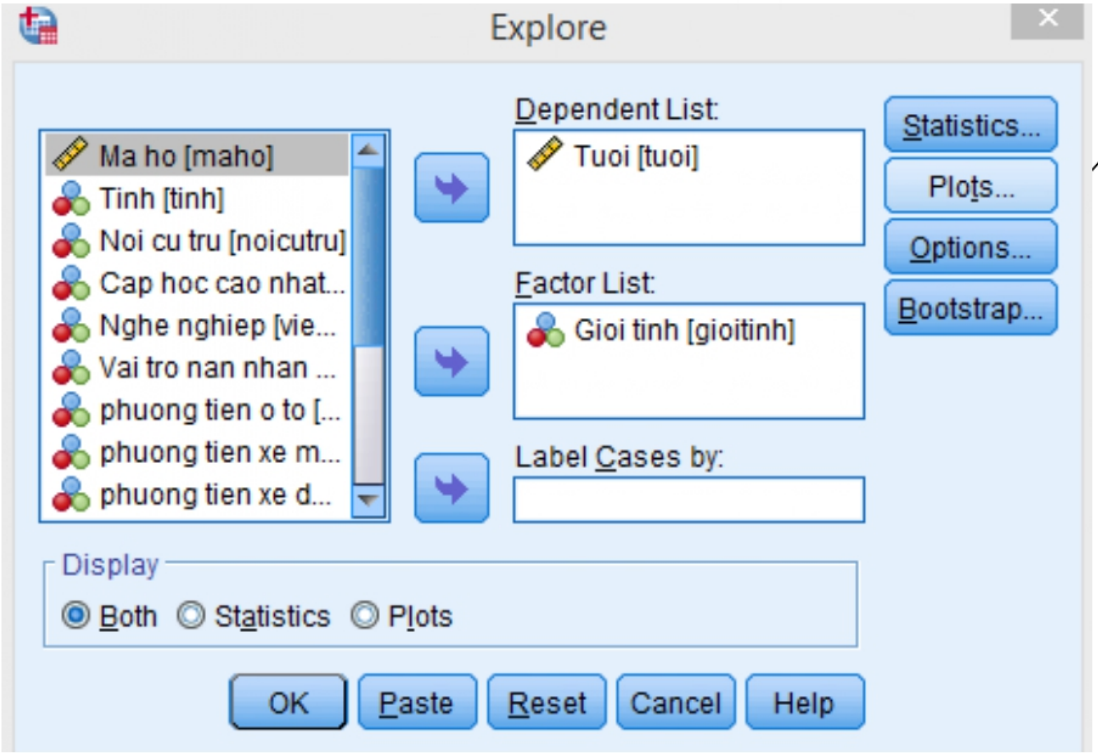


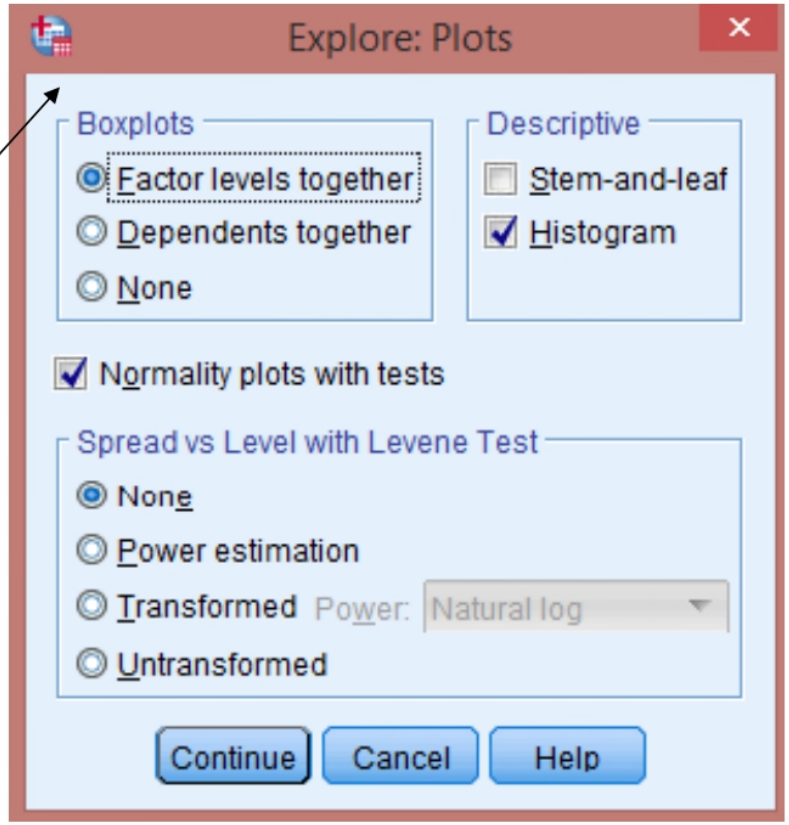


Trần Thúy Hiền

4.4. Khám phá dữ liệu (Explore)

2. Thao tác thực hiện





Trần Thúy Hiền

4.4. Khám phá dữ liệu (Explore)

Descriptives

	Gioi tinh			Statistic	Std. Error
Tuoi	Nam	Mean		34.79	3.109
		95% Confidence Interval for Mean	Lower Bound	28.41	
			Upper Bound	41.17	
		5% Trimmed Mean		34.13	
		Median		28.50	
		Variance		270.693	
		Std. Deviation		16.453	
		Minimum		14	
		Maximum		67	
		Range		53	
		Interquartile Range		25	
		Skewness		.696	.441
		Kurtosis		-.674	.858
	Nu	Mean		38.36	3.480
		95% Confidence Interval for Mean	Lower Bound	31.13	
			Upper Bound	45.60	
		5% Trimmed Mean		38.62	
		Median		42.50	
		Variance		266.433	
		Std. Deviation		16.323	
		Minimum		11	
		Maximum		61	
		Range		50	
		Interquartile Range		31	
		Skewness		-.263	.491
		Kurtosis		-1.473	.953

Trần Thúy Hiền

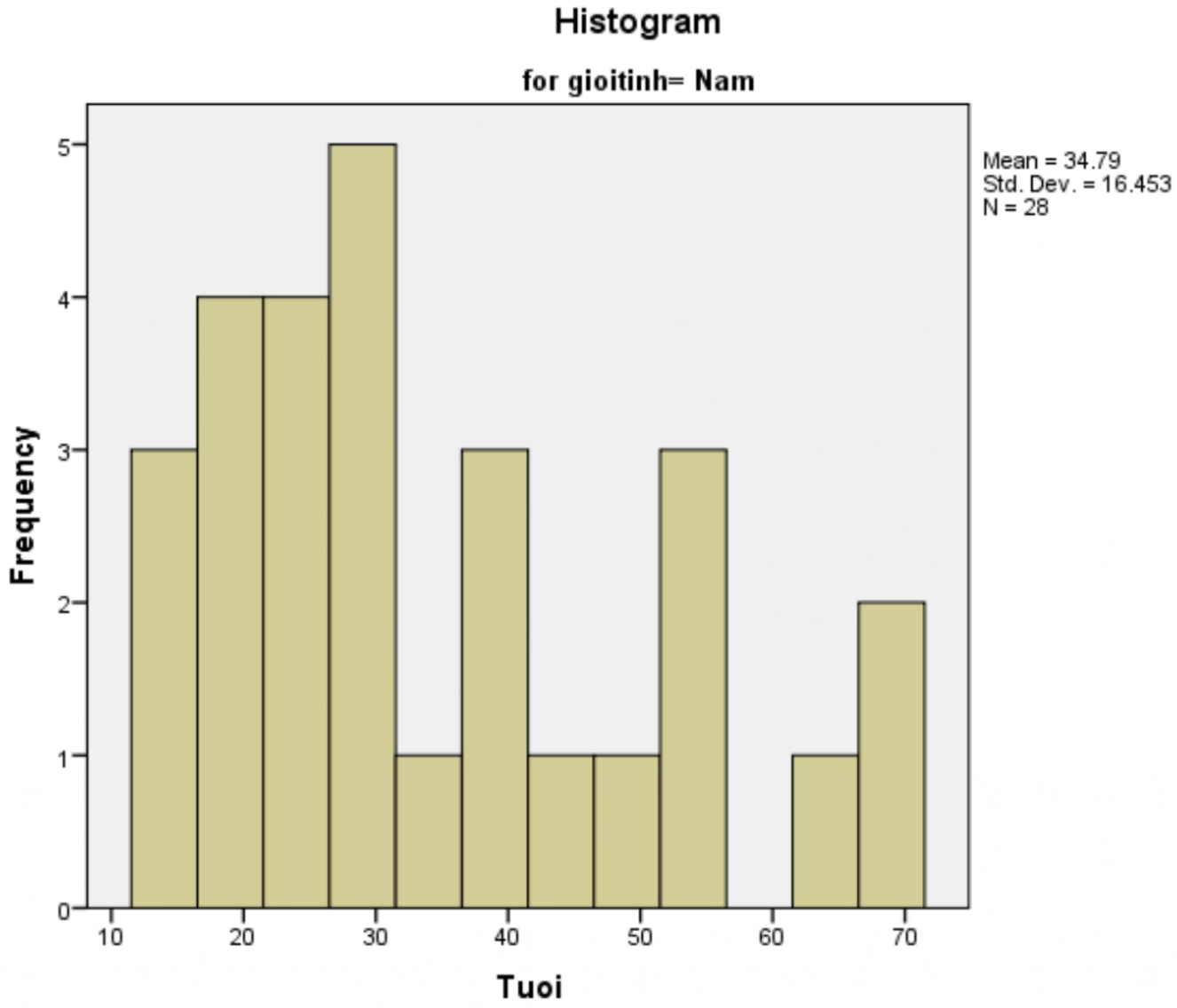
4.4. Khám phá dữ liệu (Explore)

Tests of Normality

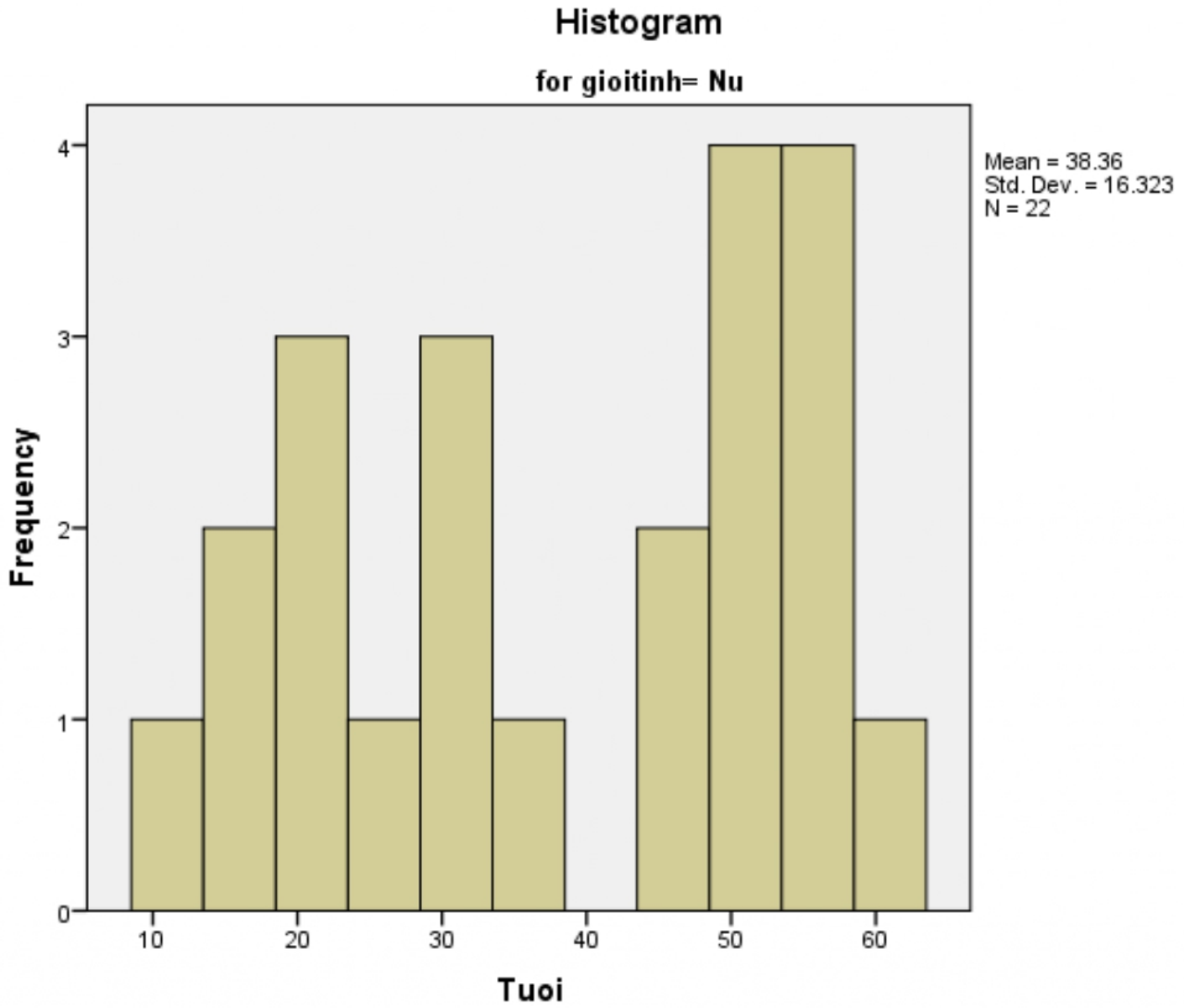
		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Gioi tinh	Statistic	df	Sig.	Statistic	df	Sig.
Tuoi	Nam	.173	28	.031	.907	28	.017
	Nu	.202	22	.020	.910	22	.047

a. Lilliefors Significance Correction

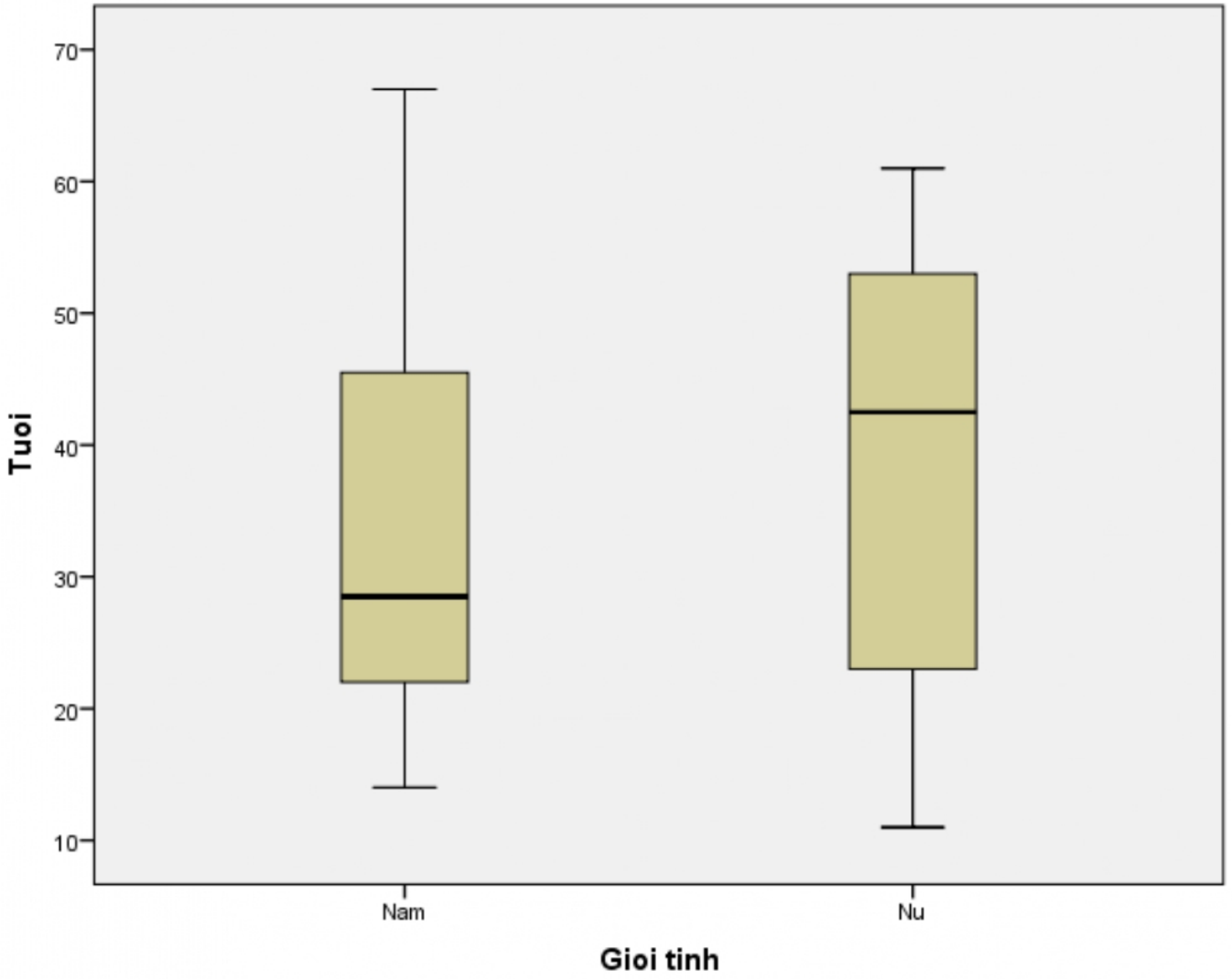
4.4. Khám phá dữ liệu (Explore)



4.4. Khám phá dữ liệu (Explore)

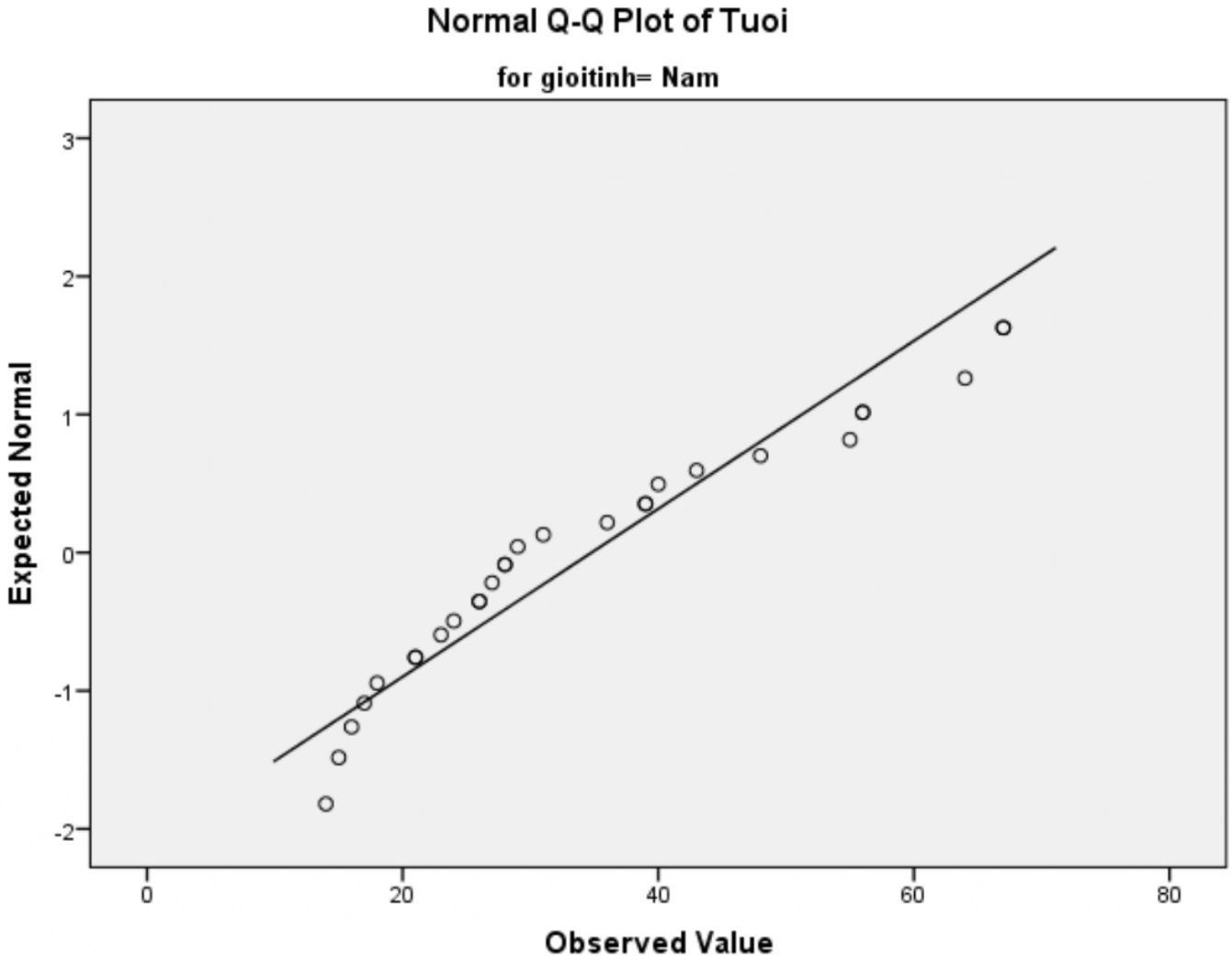


4.4. Khám phá dữ liệu (Explore)



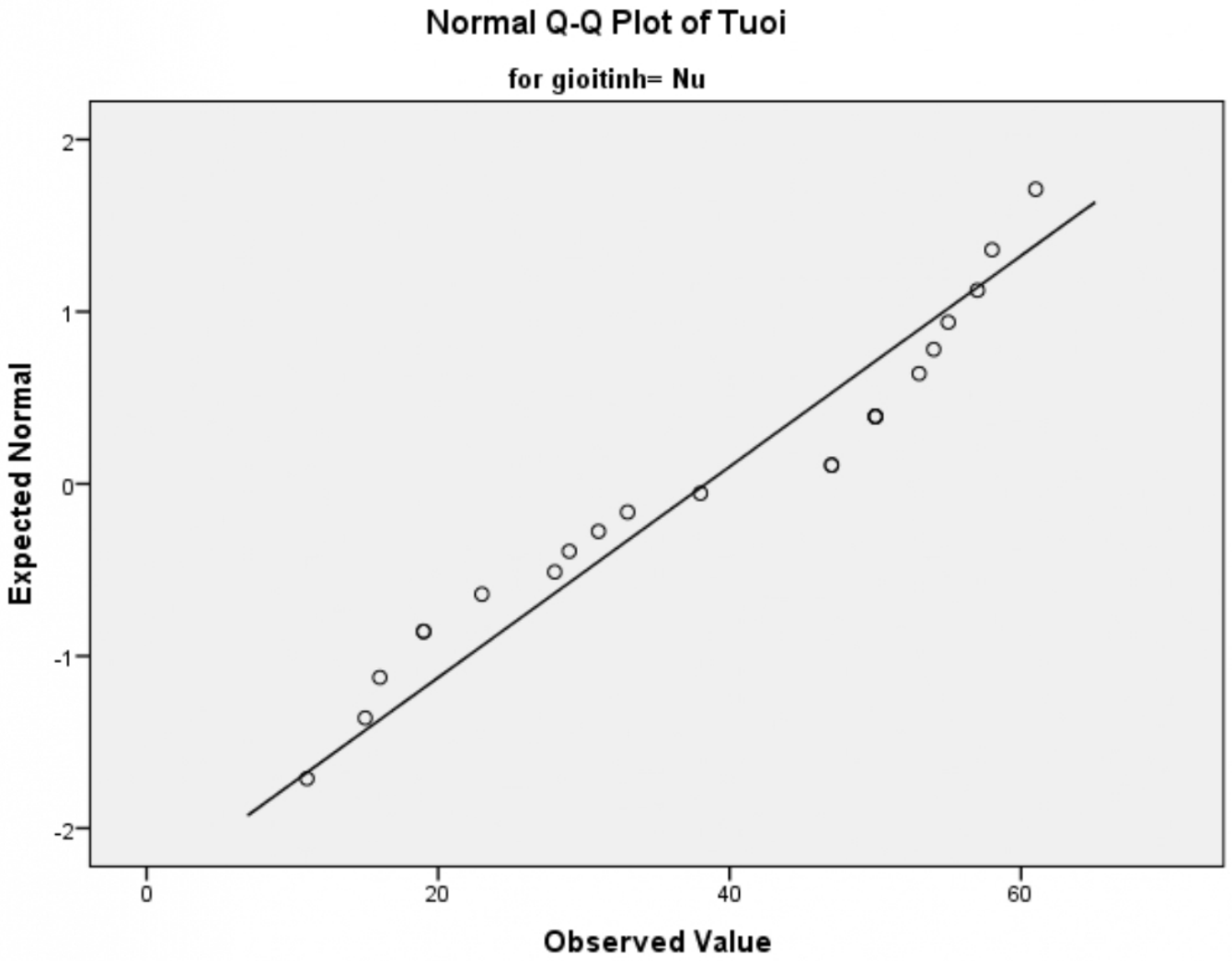
Trần Thúy Hiền

4.4. Khám phá dữ liệu (Explore)



Trần Thúy Hiền

4.4. Khám phá dữ liệu (Explore)



Trần Thúy Hiền

4.5. Bảng chéo (Crosstabs)

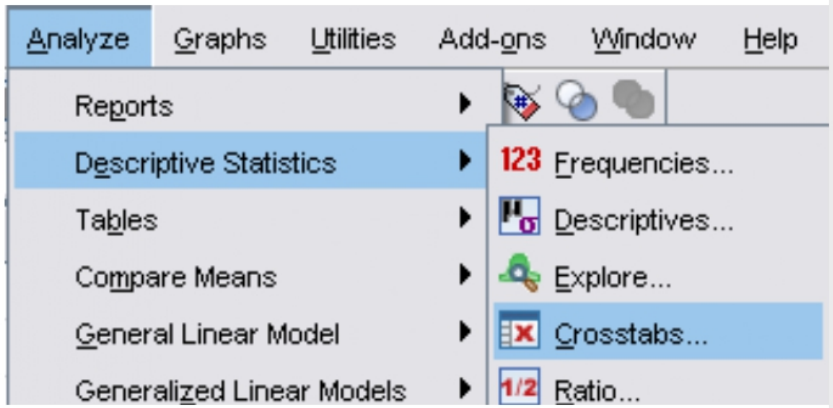
1. Chức năng

- Thể hiện tần suất xuất hiện của 1 biến này trong mối quan hệ với 1 hoặc nhiều biến khác;
- Giúp phát hiện những sai sót trong dữ liệu từ việc phát hiện ra những mối quan hệ vô lý và bất thường giữa 2 biến.

4.5. Bảng chéo (Crosstabs)

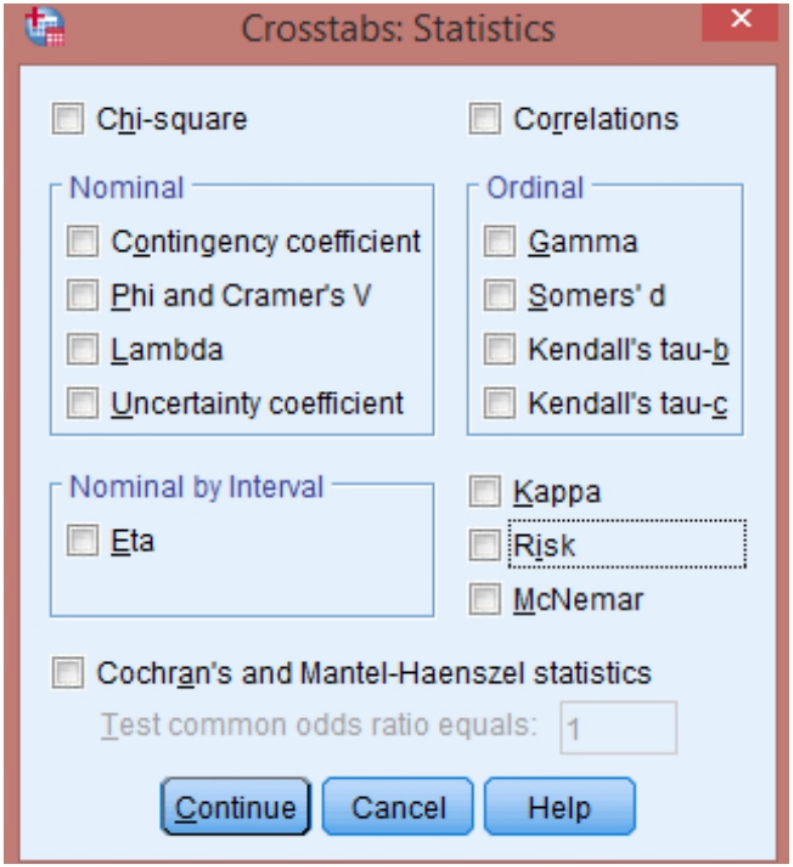
2. Thao tác thực hiện

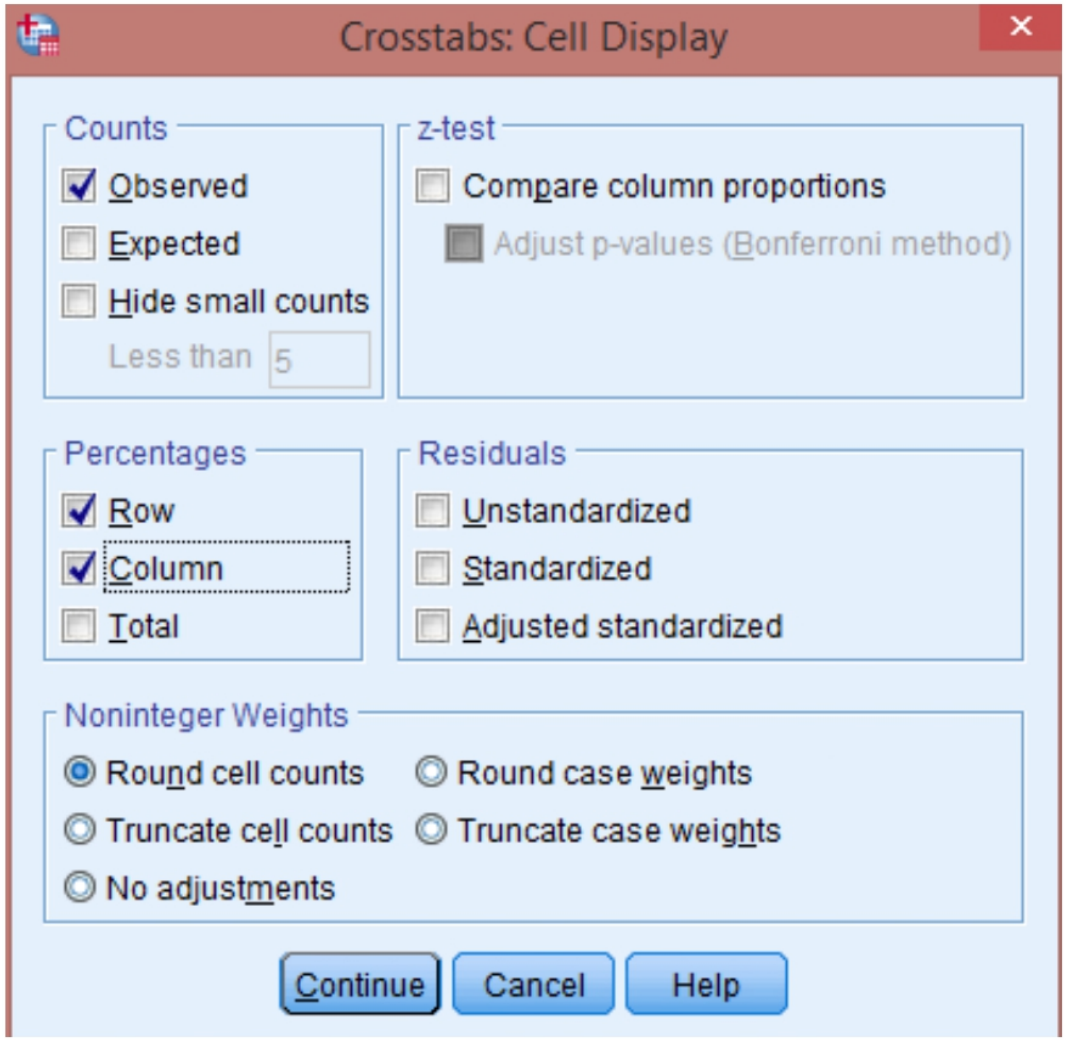
- Analyze/Descriptive Statistics/Crosstabs



Trần Thúy Hiền

4.5. Bảng chéo (Crosstabs)





Trần Thúy Hiền

4.5. Bảng chéo (Crosstabs)

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Gioi tinh * Tinh	50	100.0%	0	0.0%	50	100.0%

Gioi tinh * Tinh Crosstabulation

			Tinh			Total
			Quang Binh	Quang Tri	Thua Thien Hue	
Gioi tinh	Nam	Count	10	9	9	28
		% within Gioi tinh	35.7%	32.1%	32.1%	100.0%
		% within Tinh	55.6%	45.0%	75.0%	56.0%
	Nu	Count	8	11	3	22
		% within Gioi tinh	36.4%	50.0%	13.6%	100.0%
		% within Tinh	44.4%	55.0%	25.0%	44.0%
Total		Count	18	20	12	50
		% within Gioi tinh	36.0%	40.0%	24.0%	100.0%
		% within Tinh	100.0%	100.0%	100.0%	100.0%

4.6. Bảng tùy chỉnh (Custom Tables)

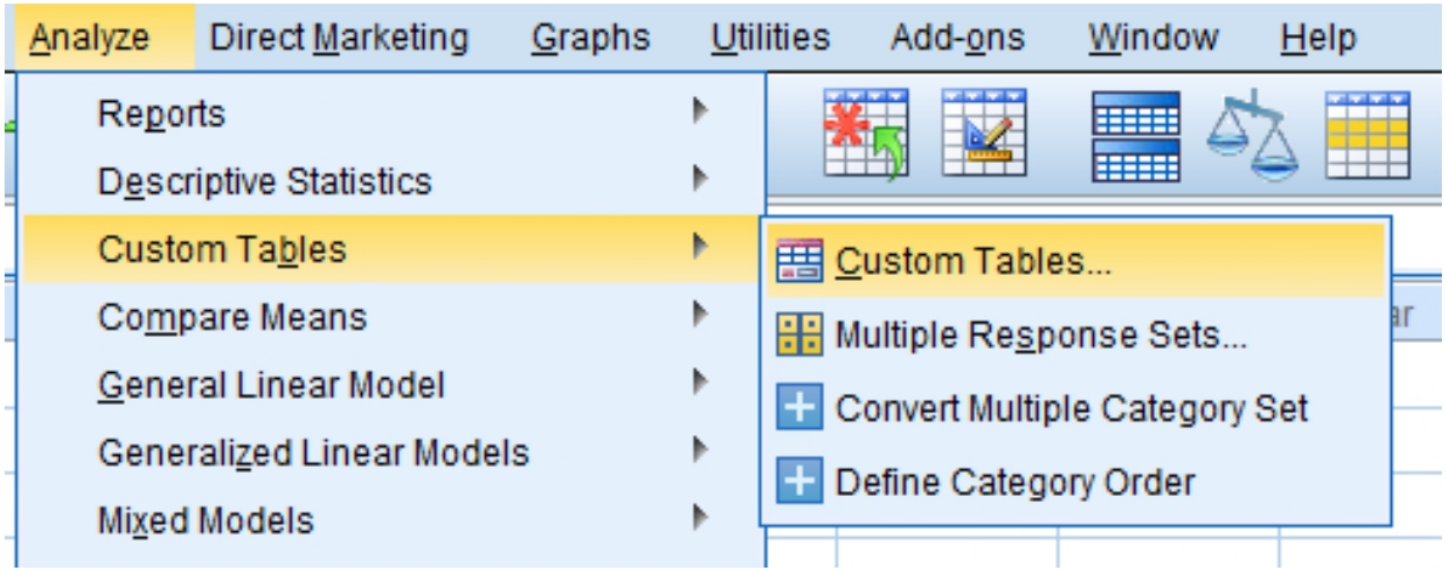
1. Chức năng

- Thể hiện tần suất xuất hiện của 1 biến này trong mối quan hệ với 1 hay nhiều biến khác;
- Tương tự bảng Crosstabs, nhưng có nhiều tính năng, tùy chọn về mô tả dữ liệu hơn.

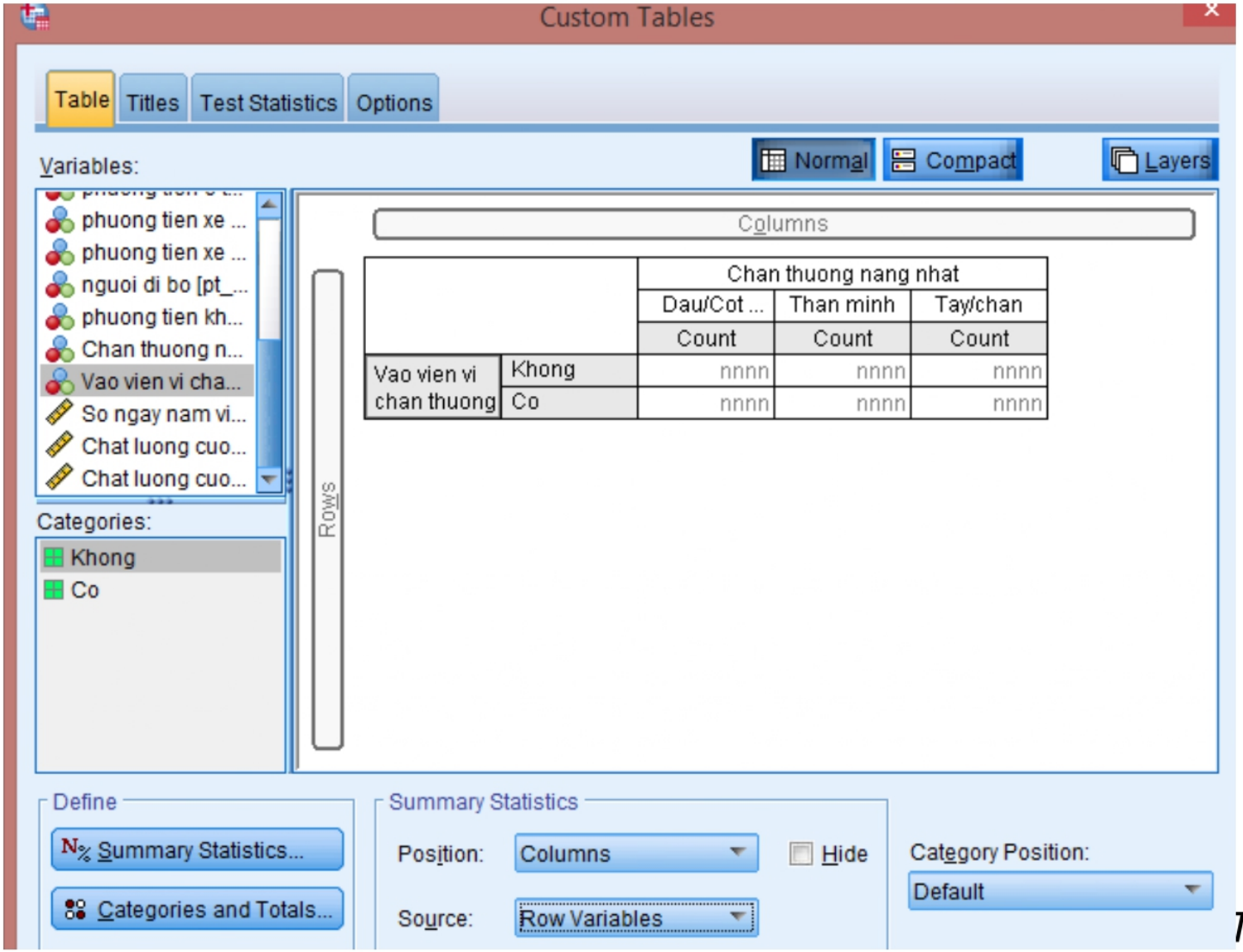
4.6. Bảng tùy chỉnh (Custom Tables)

2. Thao tác thực hiện

- Analyze/Custom Tables/Custom Tables

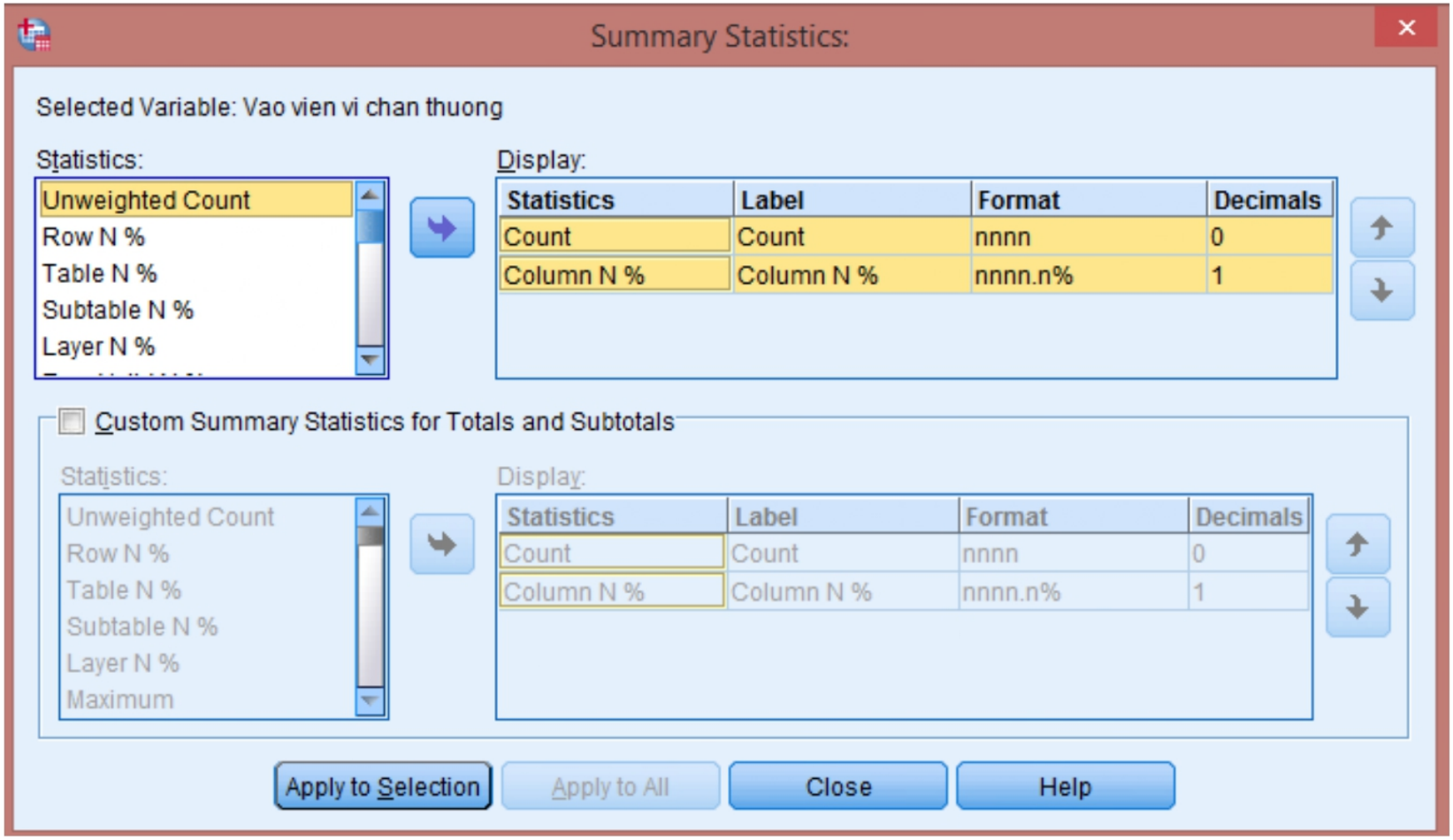


4.6. Bảng tùy chỉnh (Custom Tables)



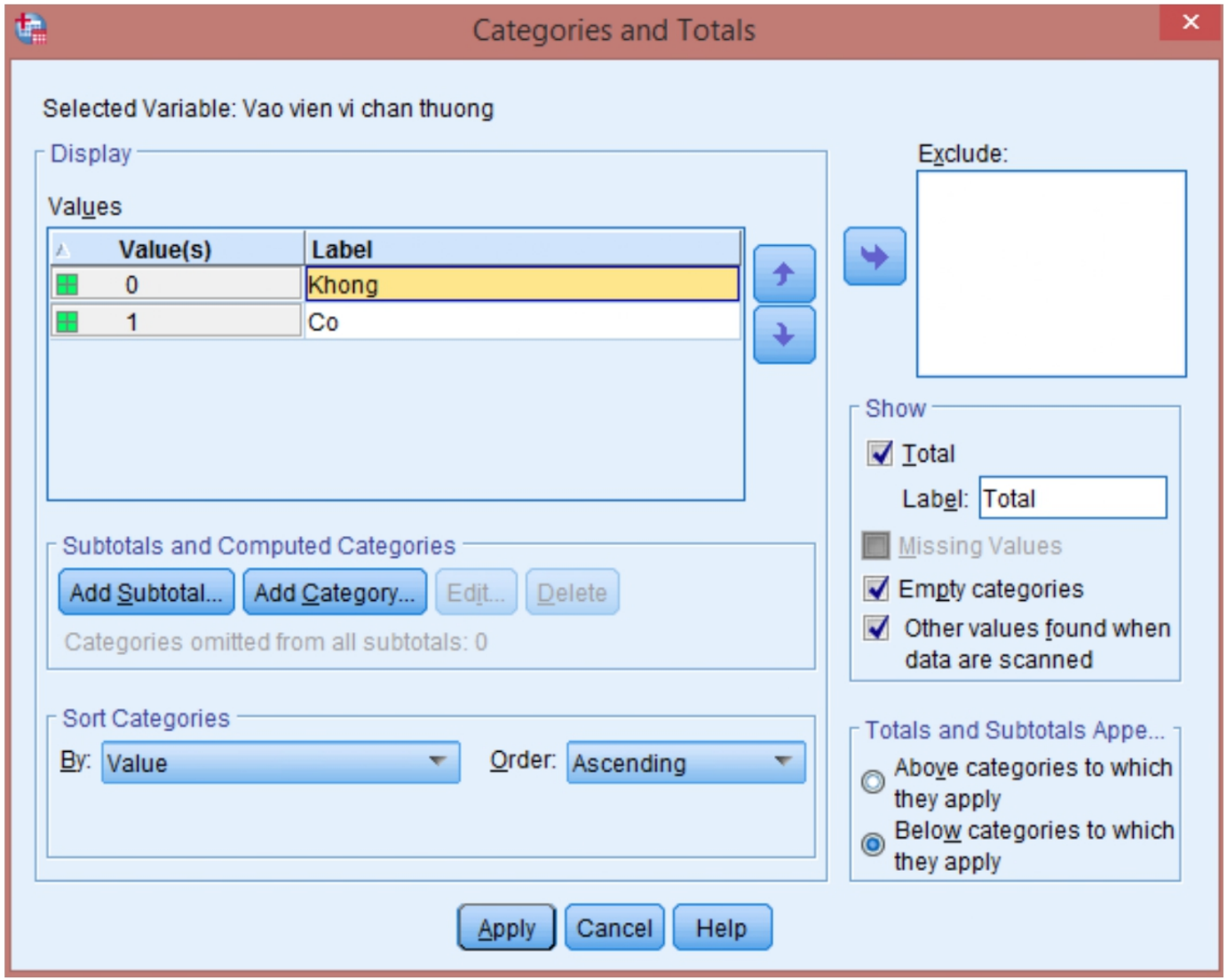
Trần Thúy Hiền

4.6. Bảng tùy chỉnh (Custom Tables)



Trần Thúy Hiền

4.6. Bảng tùy chỉnh (Custom Tables)



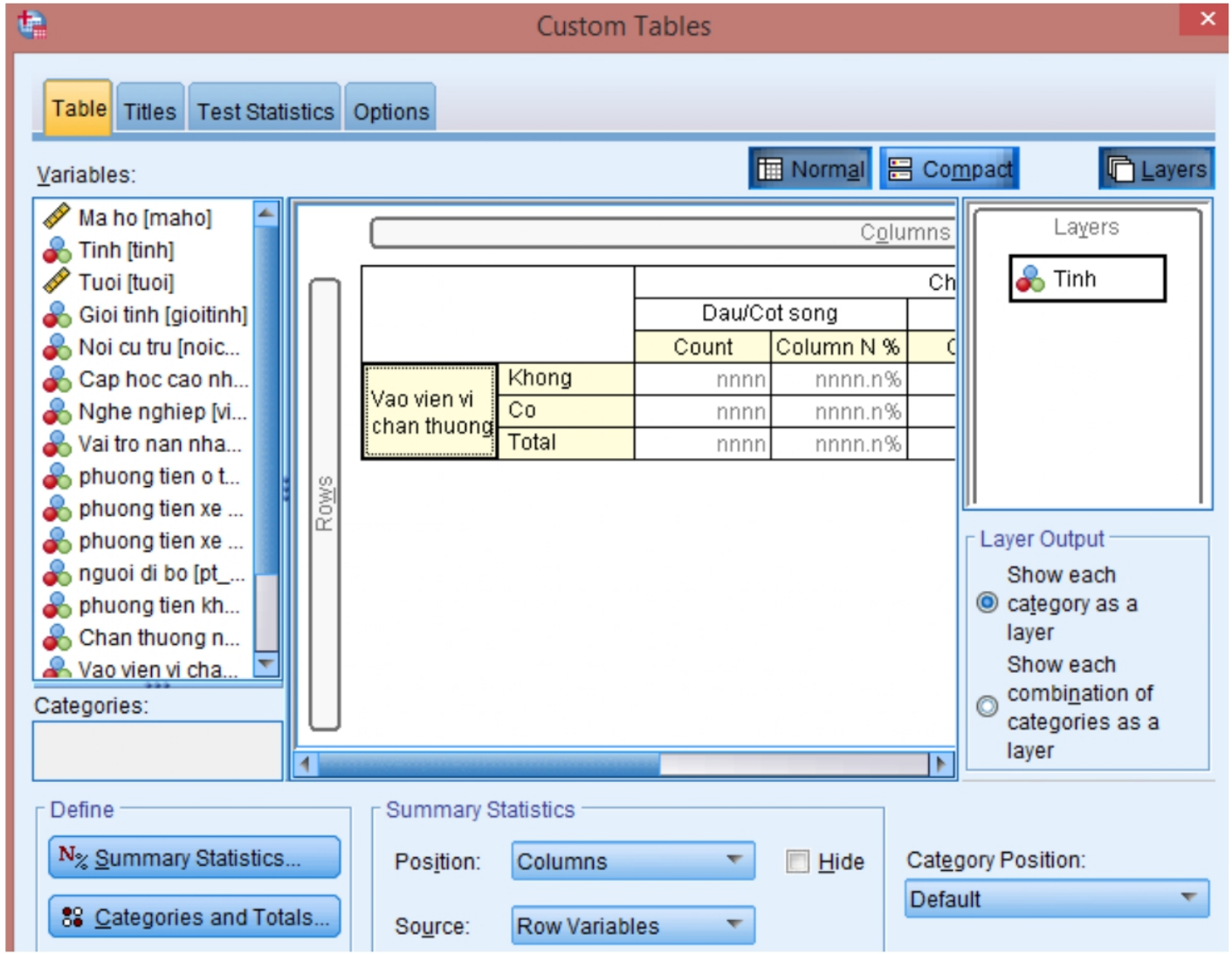
Trần Thúy Hiền

4.6. Bảng tùy chỉnh (Custom Tables)

		Chan thuong nang nhat					
		Dau/Cot song		Than minh		Tay/chan	
		Count	Column N %	Count	Column N %	Count	Column N %
Vao vien vi chan thuong	Khong	7	41.2%	8	44.4%	8	53.3%
	Co	10	58.8%	10	55.6%	7	46.7%
	Total	17	100.0%	18	100.0%	15	100.0%

Trần Thúy Hiền

4.6. Bảng tùy chỉnh (Custom Tables)



Trần Thúy Hiền

4.6. Bảng tùy chỉnh (Custom Tables)

Tinh Quang Binh

		Chan thuong nang nhat					
		Dau/Cot song		Than minh		Tay/chan	
		Count	Column N %	Count	Column N %	Count	Column N %
Vao vien vi chan thuong	Khong	3	50.0%	3	50.0%	4	66.7%
	Co	3	50.0%	3	50.0%	2	33.3%
	Total	6	100.0%	6	100.0%	6	100.0%

Custom Tables

Layer Tinh Quang Binh

- Tinh Quang Binh
- Tinh Quang Tri
- Tinh Thua Thien Hue

		Chan thuong nang nhat					
		Dau/Cot song		Than minh		Tay/chan	
		Count	Column N %	Count	Column N %	Count	Column N %
Vao vien vi chan thuong	Khong	3	50.0%	3	50.0%	4	66.7%
	Co	3	50.0%	3	50.0%	2	33.3%
	Total	6	100.0%	6	100.0%	6	100.0%

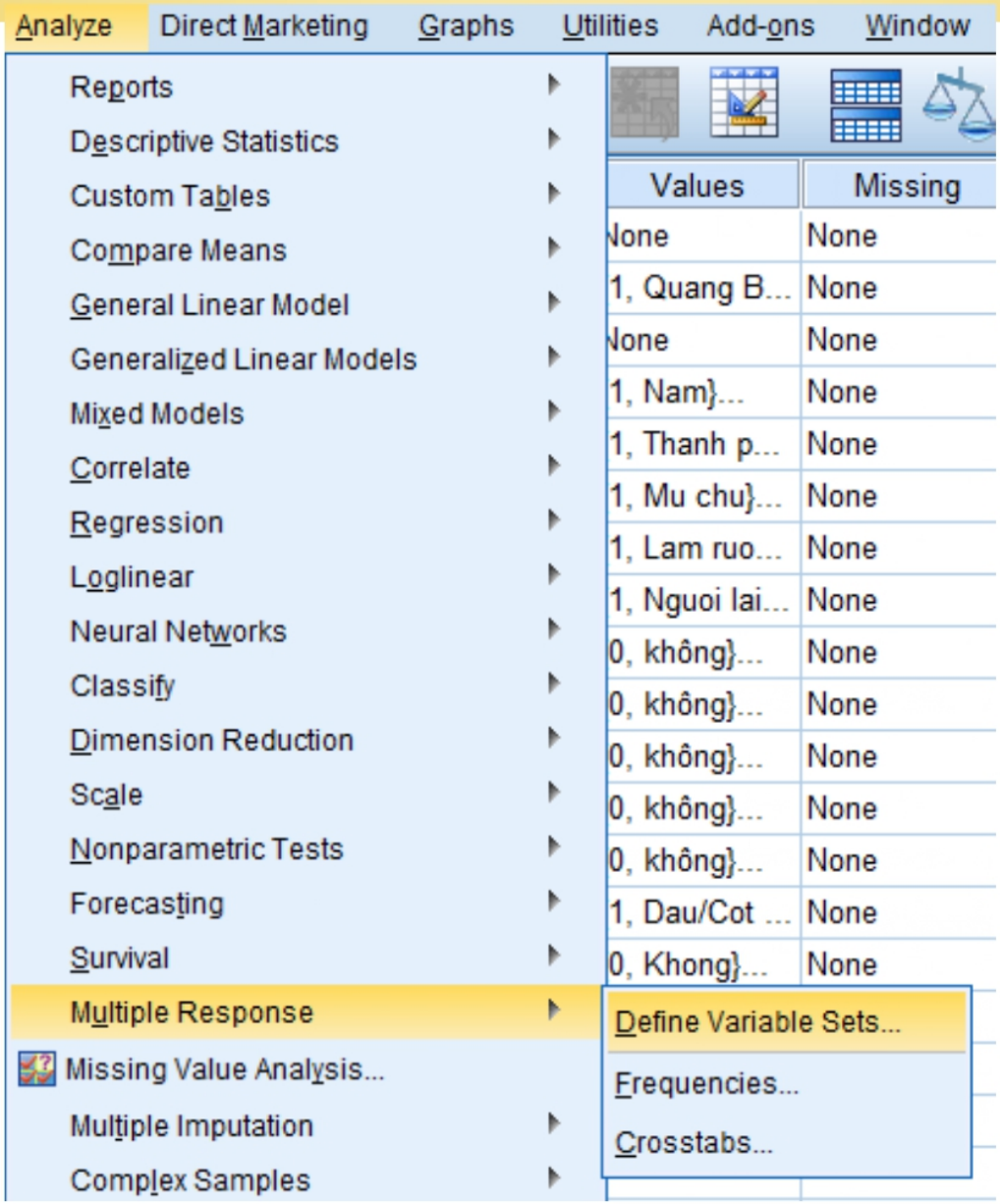
Trần Thúy Hiền

4.7. Xử lý câu hỏi nhiều trả lời

1. Định nghĩa nhóm biến nhiều trả lời

- Câu hỏi nhiều trả lời: gồm nhiều biến chứa đựng các trả lời có thể có, được gọi là các biến sơ cấp;
- Trước khi thực hiện các phân tích thống kê liên quan đến câu hỏi nhiều trả lời: phải gộp các biến sơ cấp thành một **biến gộp**;
- **Biến gộp** chứa toàn bộ các giá trị trong các biến sơ cấp của một câu hỏi nhiều trả lời.

4.7. Xử lý câu hỏi nhiều trả lời

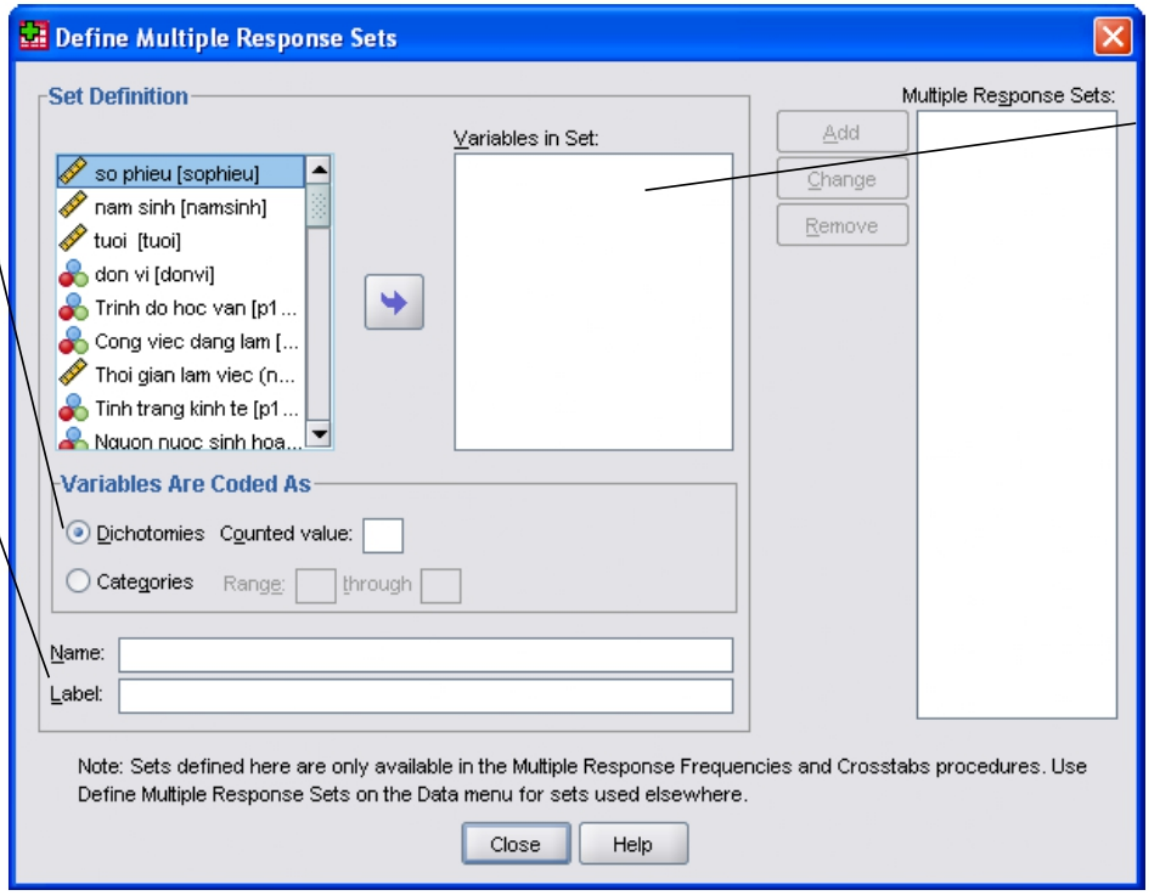


Trần Thúy Hiền

4.7. Xử lý câu hỏi nhiều trả lời

Cách mã hóa các biến

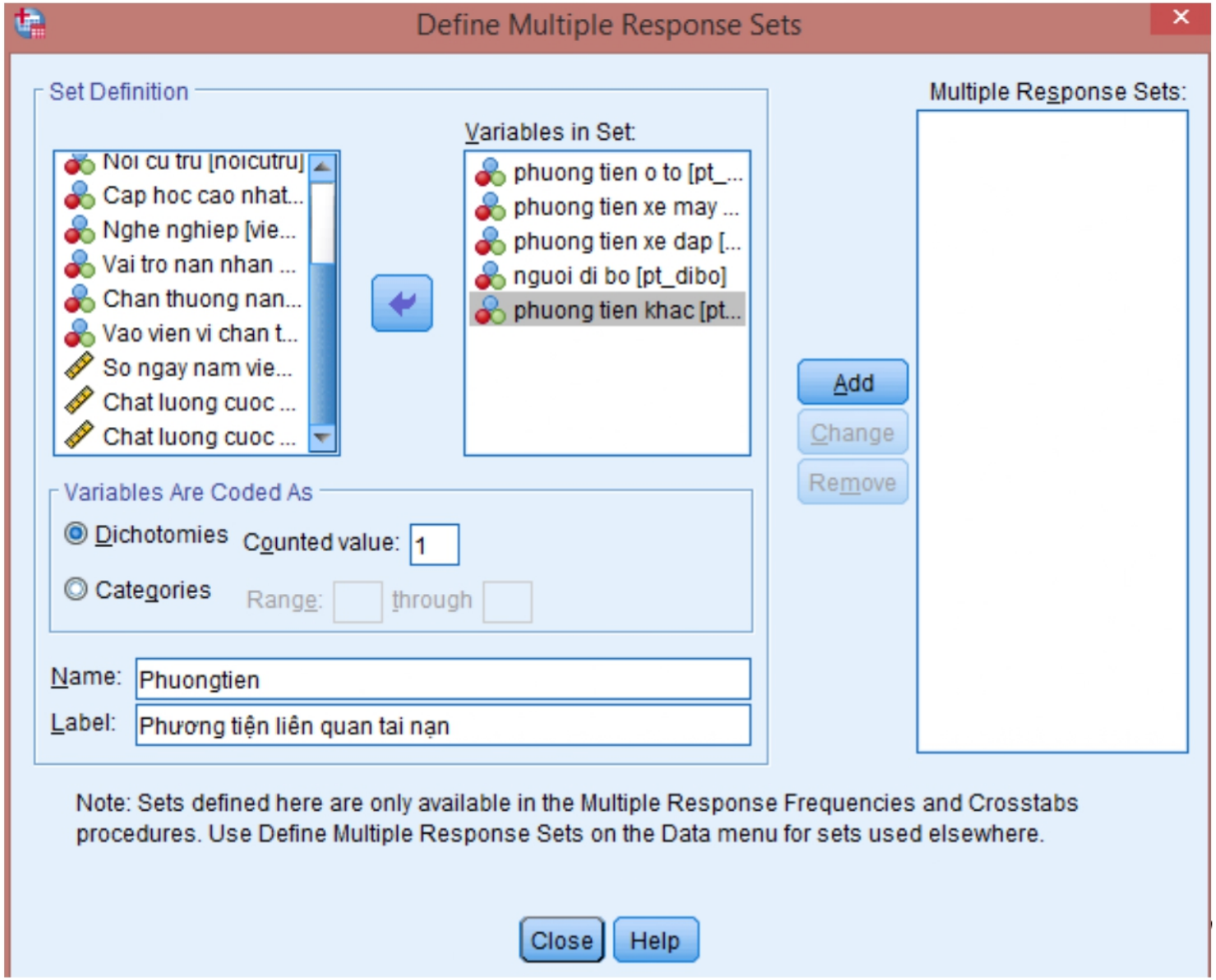
Tên và nhãn của biến gộp



Nhóm biến sơ cấp của 1 câu hỏi nhiều trả lời

Trần Thúy Hiền

4.7. Xử lý câu hỏi nhiều trả lời



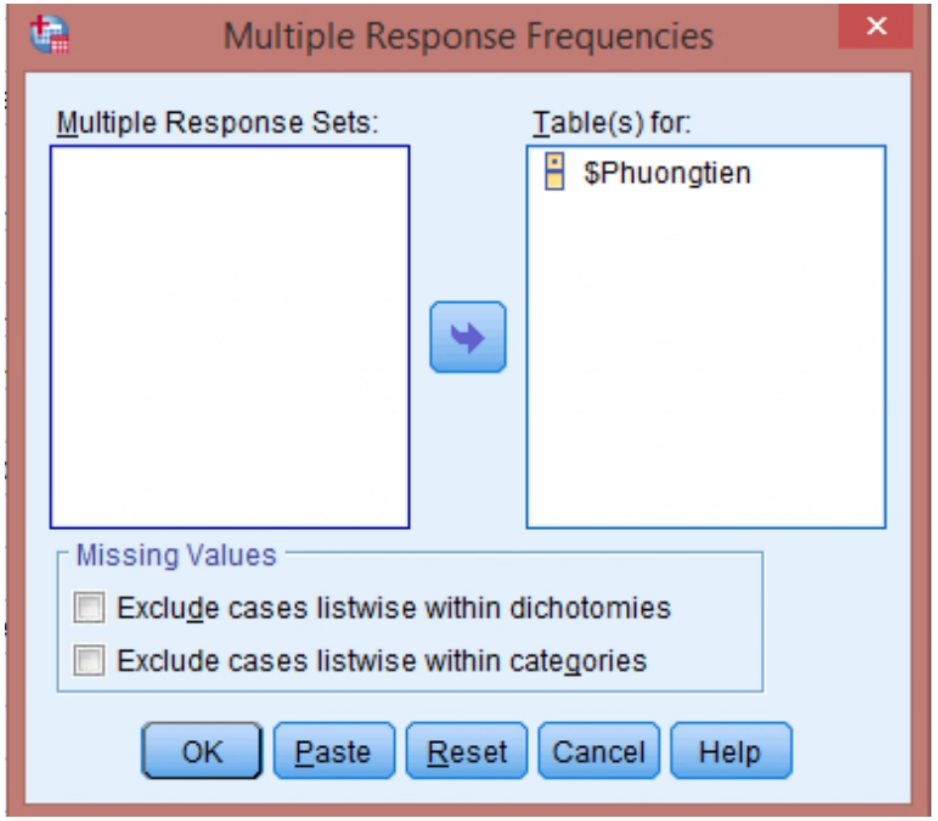
Thúy Hiền

4.7. Xử lý câu hỏi nhiều trả lời

2. Lập bảng cho biến nhiều trả lời

- Analyze/Multiple Response chọn Frequencies hoặc Crosstabs

a) Ví dụ: Lập bảng Frequencies cho biến gộp **Phuongtien**



4.7. Xử lý câu hỏi nhiều trả lời

Multiple Response

Case Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
$Phuongtien[a]	50	100.0%	0	0.0%	50	100.0%

a. Dichotomy group tabulated at value 1.

$Phuongtien Frequencies

		Responses		Percent of Cases
		N	Percent	
Phương tiện liên quan tai nạn[a]	phuong tien o to	26	25.7%	52.0%
	phuong tien xe may	30	29.7%	60.0%
	phuong tien xe dap	17	16.8%	34.0%
	nguoi di bo	14	13.9%	28.0%
	phuong tien khac	14	13.9%	28.0%
Total		101	100.0%	202.0%

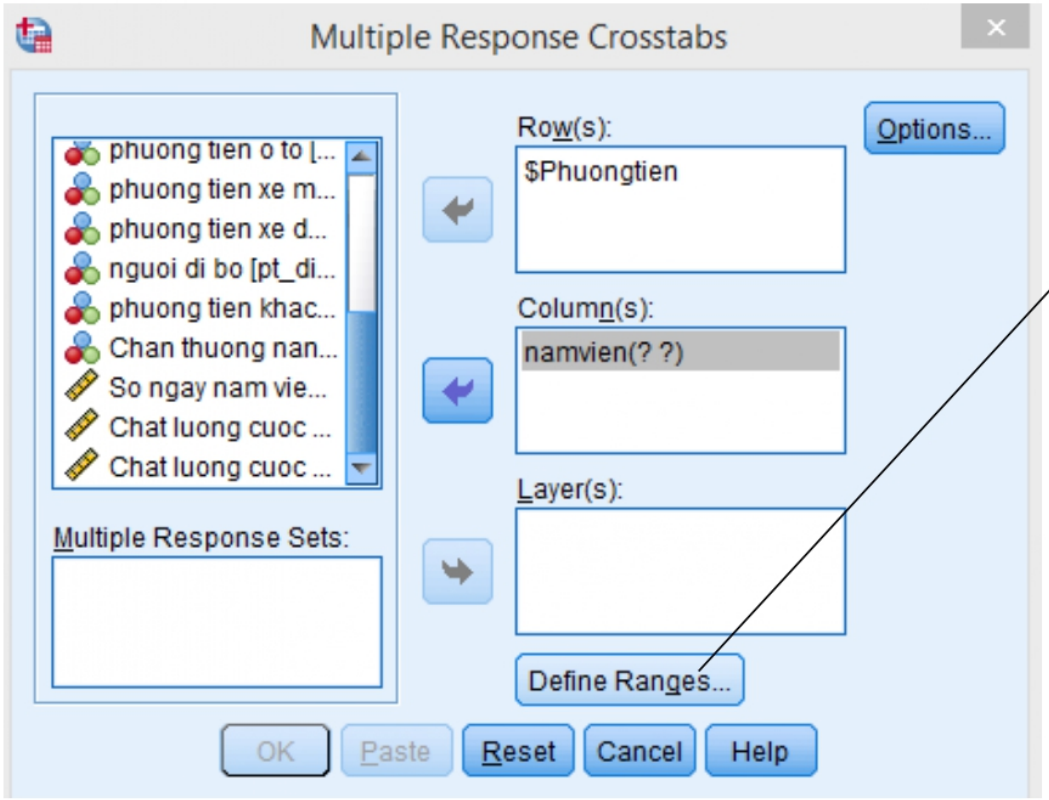
a. Dichotomy group tabulated at value 1.

Trần Thúy Hiền

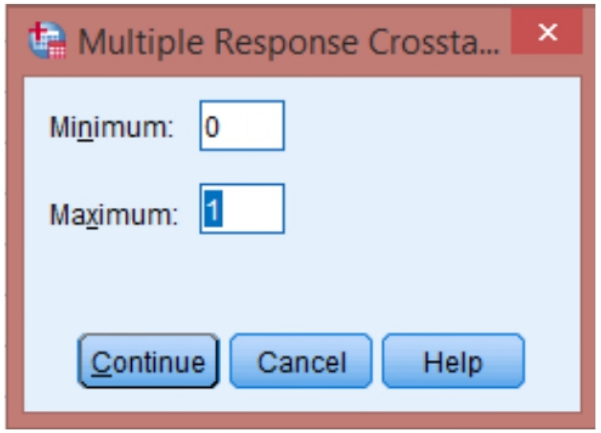
4.7. Xử lý câu hỏi nhiều trả lời

2. Lập bảng cho biến nhiều trả lời

b) Ví dụ: Lập bảng Crosstabs cho biến gộp Phuongtien và biến namvien.

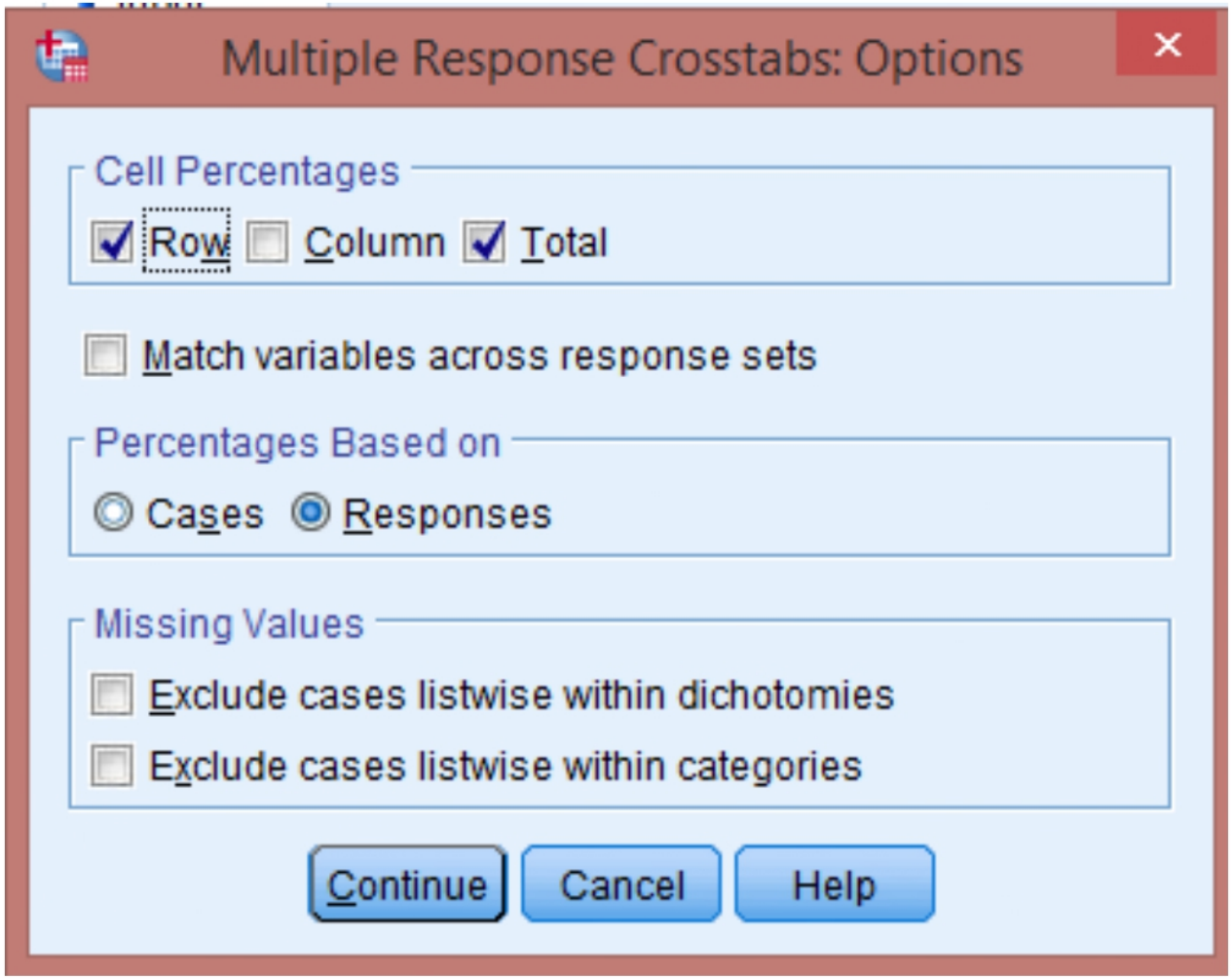


Định nghĩa dải giá trị cho biến namvien



Trần Thúy Hiền

4.7. Xử lý câu hỏi nhiều trả lời



Trần Thúy Hiền

4.7. Xử lý câu hỏi nhiều trả lời

Case Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
$Phuongtien*namvien	50	100.0%	0	0.0%	50	100.0%

$Phuongtien*namvien Crosstabulation

			Vao vien vi chan thuong		
			Khong	Co	Total
Phương tiện liên quan tai nạn[a]	phuong tien o to	Count	10	16	26
		% within $Phuongtien	38.5%	61.5%	
		% of Total	9.9%	15.8%	25.7%
	phuong tien xe may	Count	12	18	30
		% within $Phuongtien	40.0%	60.0%	
		% of Total	11.9%	17.8%	29.7%
	phuong tien xe dap	Count	8	9	17
		% within $Phuongtien	47.1%	52.9%	
		% of Total	7.9%	8.9%	16.8%
	nguoi di bo	Count	9	5	14
		% within $Phuongtien	64.3%	35.7%	
		% of Total	8.9%	5.0%	13.9%
	phuong tien khac	Count	7	7	14
		% within $Phuongtien	50.0%	50.0%	
		% of Total	6.9%	6.9%	13.9%
Total		Count	46	55	101
		% of Total	45.5%	54.5%	100.0%

Percentages and totals are based on responses.

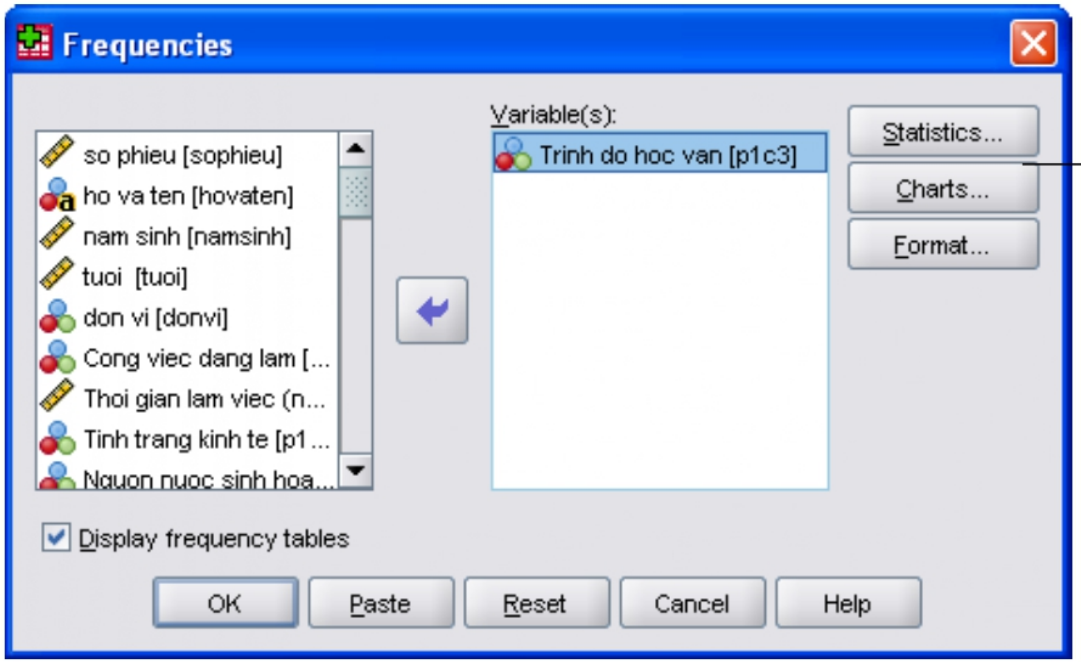
a. Dichotomy group tabulated at value 1.

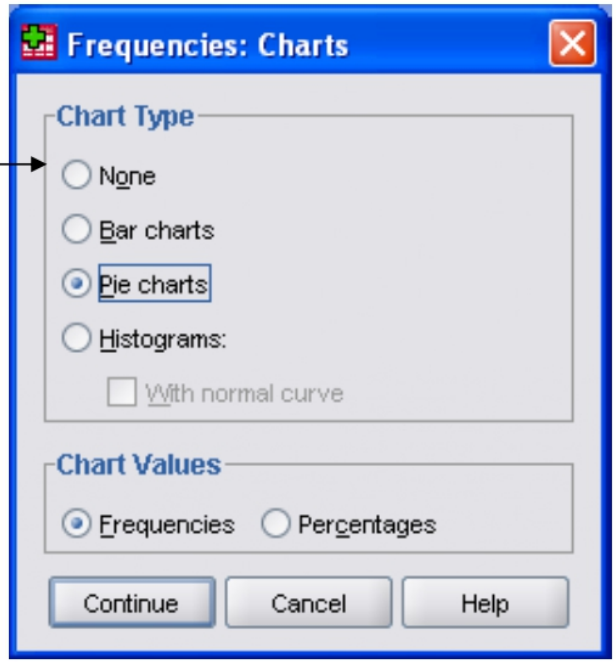
Hiền

4.8. Trình bày kết quả bằng đồ thị

1. Dữ liệu đơn giản

- Analyze/Descriptive Statistics/Frequencies
- Click nút Charts… chọn biểu đồ thích hợp



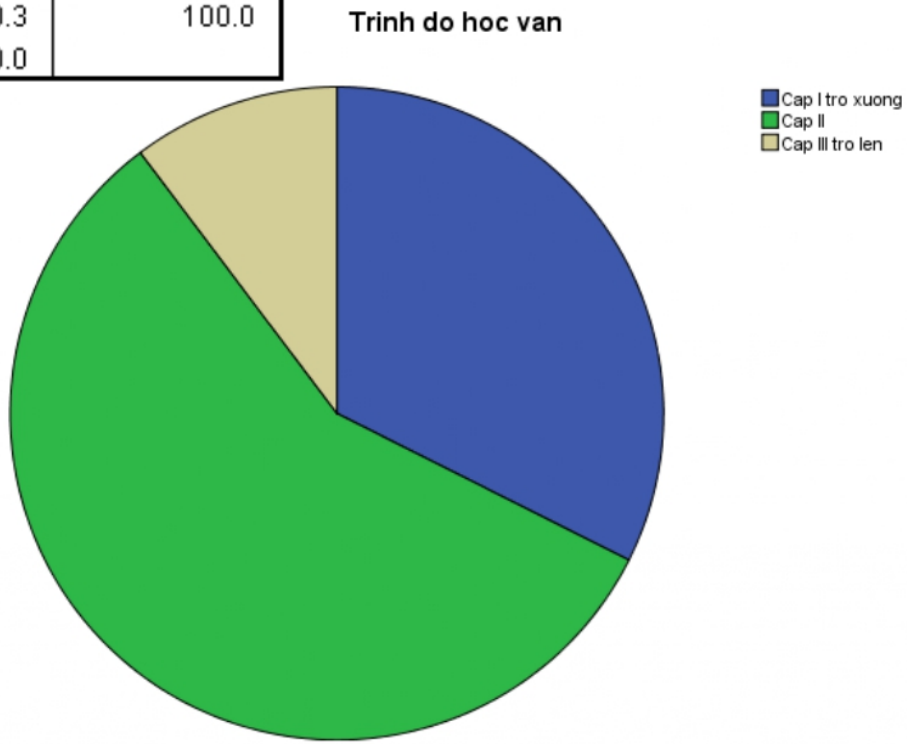


Trần Thúy Hiền

4.8. Trình bày kết quả bằng đồ thị

Trinh do hoc van

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Cap I tro xuong	164	32.4	32.4	32.4
	Cap II	290	57.3	57.3	89.7
	Cap III tro len	52	10.3	10.3	100.0
	Total	506	100.0	100.0	

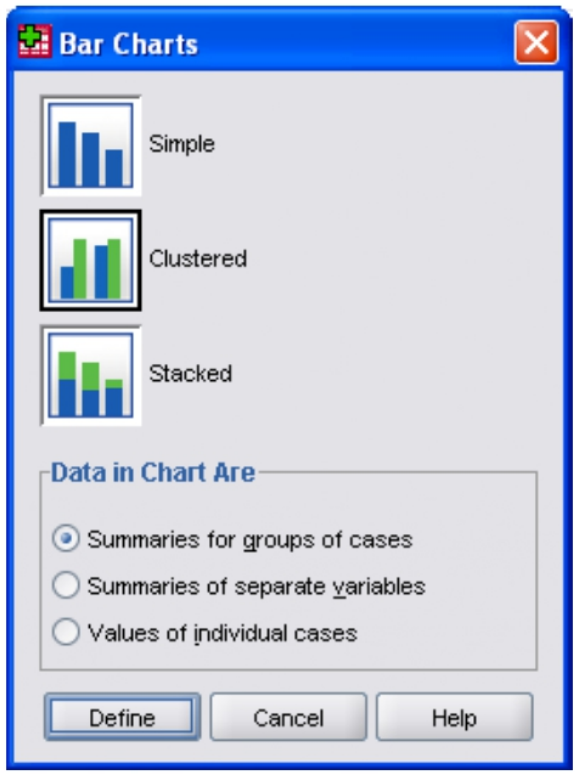


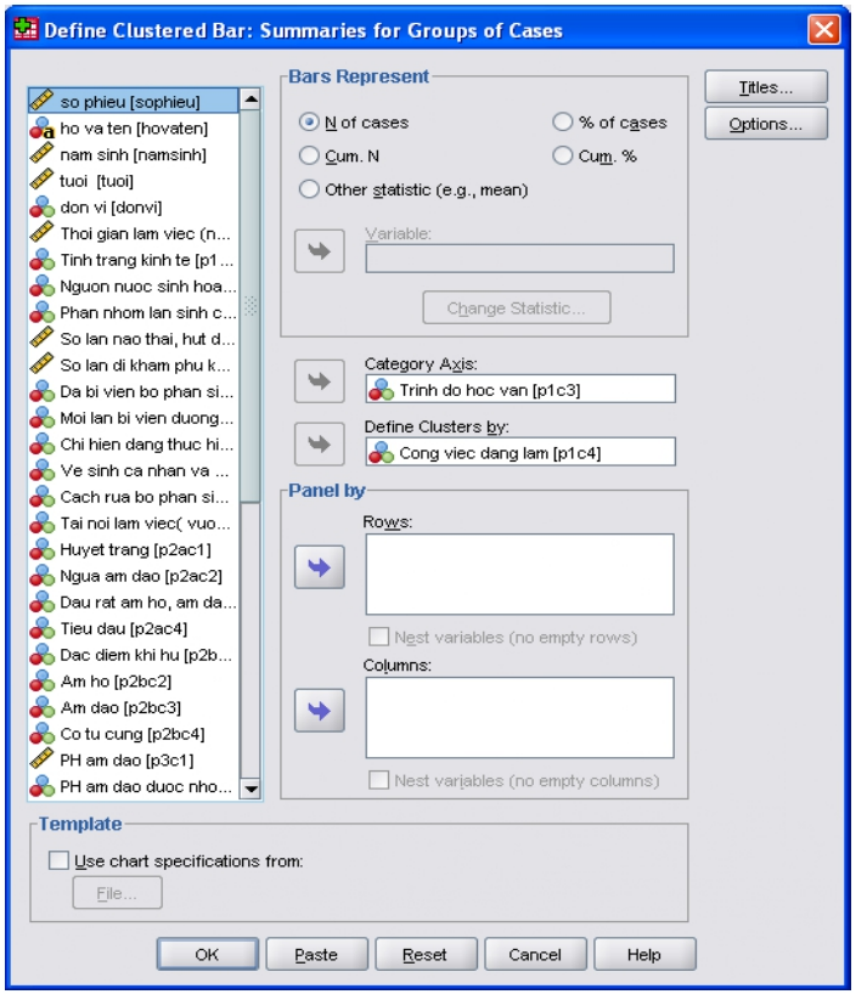
4.8. Trình bày kết quả bằng đồ thị

2. Biểu diễn dữ liệu của biến theo tính chất

- Ví dụ: Biểu diễn trình độ học vấn của đối tượng theo công việc đang làm
  - Graphs/Legacy Dialog/Bar…
  - Chọn nút Clustered
  - Click nút Define

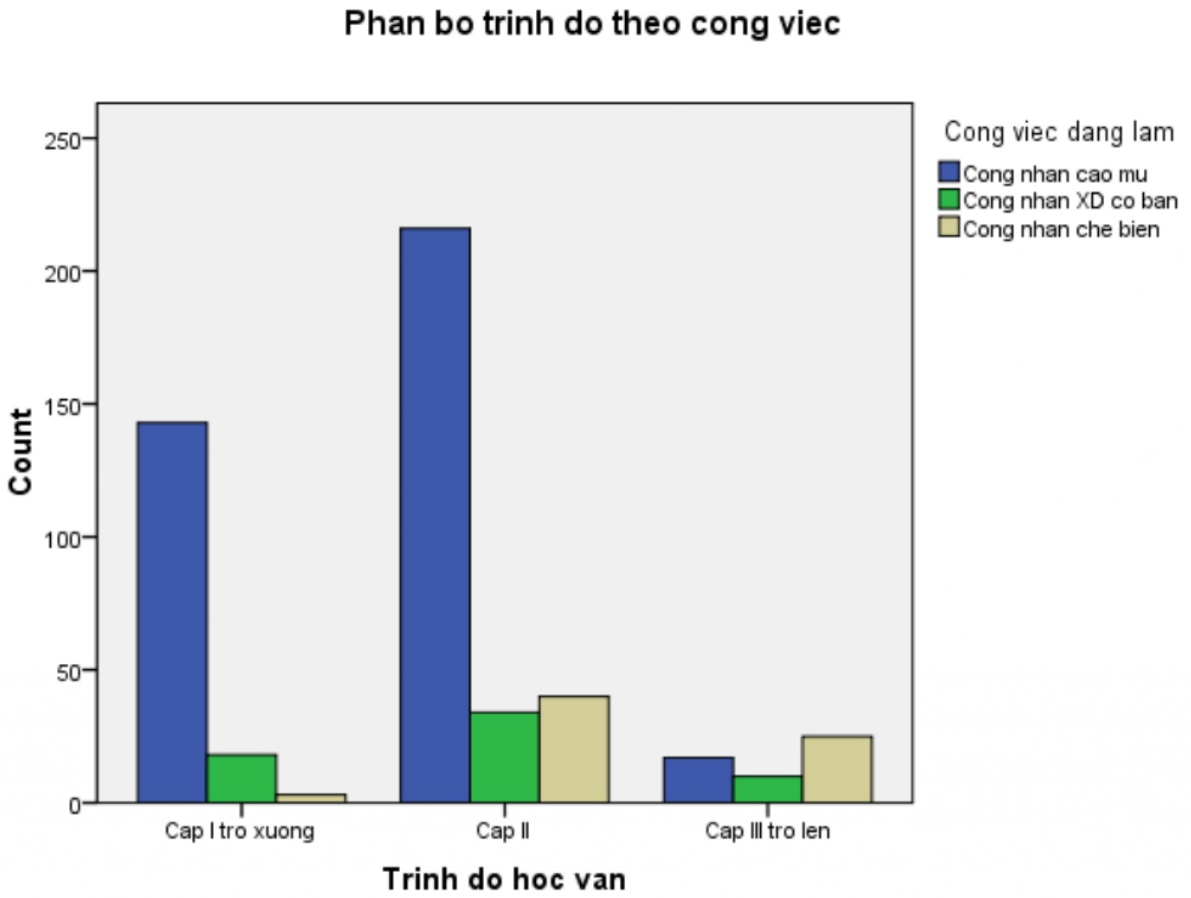
4.8. Trình bày kết quả bằng đồ thị





Trần Thúy Hiền

4.8. Trình bày kết quả bằng đồ thị



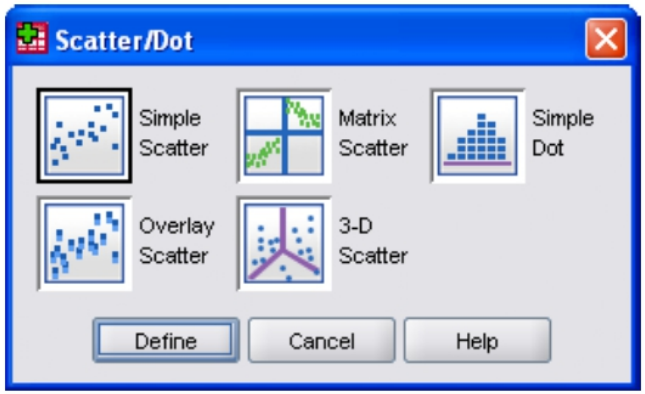
Trần Thúy Hiền

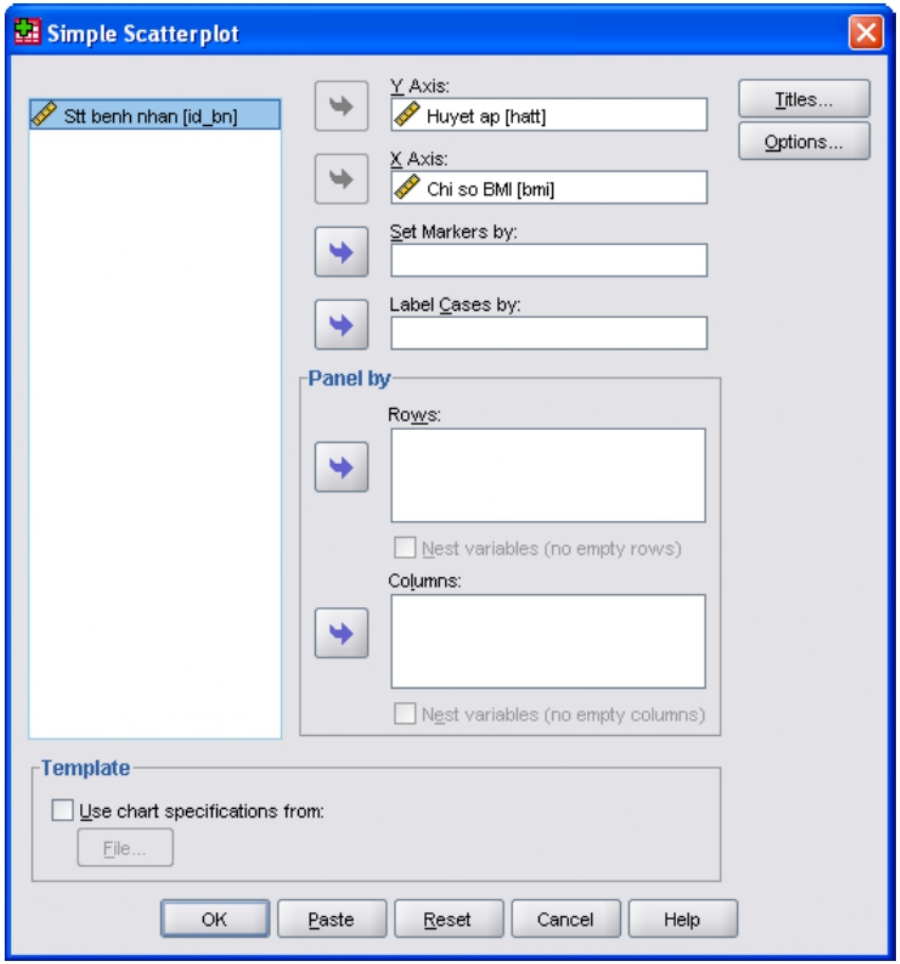
4.8. Trình bày kết quả bằng đồ thị

3. Vẽ đồ thị tương quan

- Sử dụng cho 2 biến định lượng liên tục có tương quan
- Vd: Biểu diễn mối tương quan giữa chỉ số BMI và huyết áp của đối tượng nghiên cứu
- Graphs/Legacy Dialog chọn Scatter/Dot…
- Chọn Simple Scatter
- Click Define

4.8. Trình bày kết quả bằng đồ thị

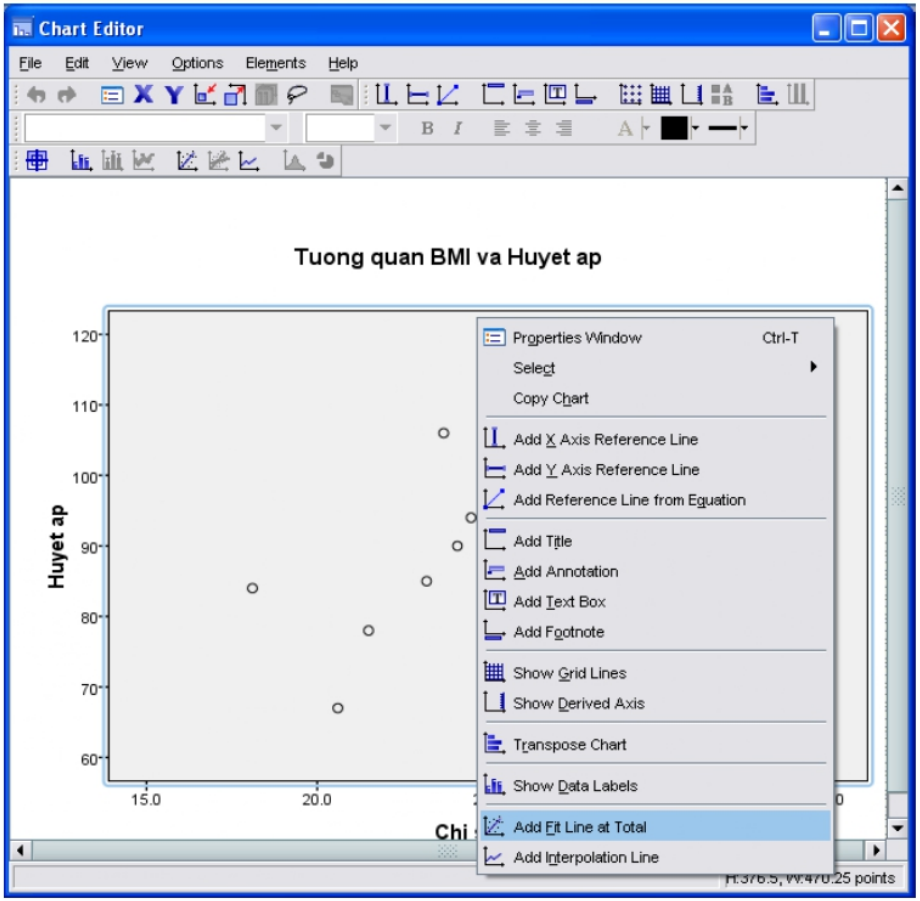




Trần Thúy Hiền

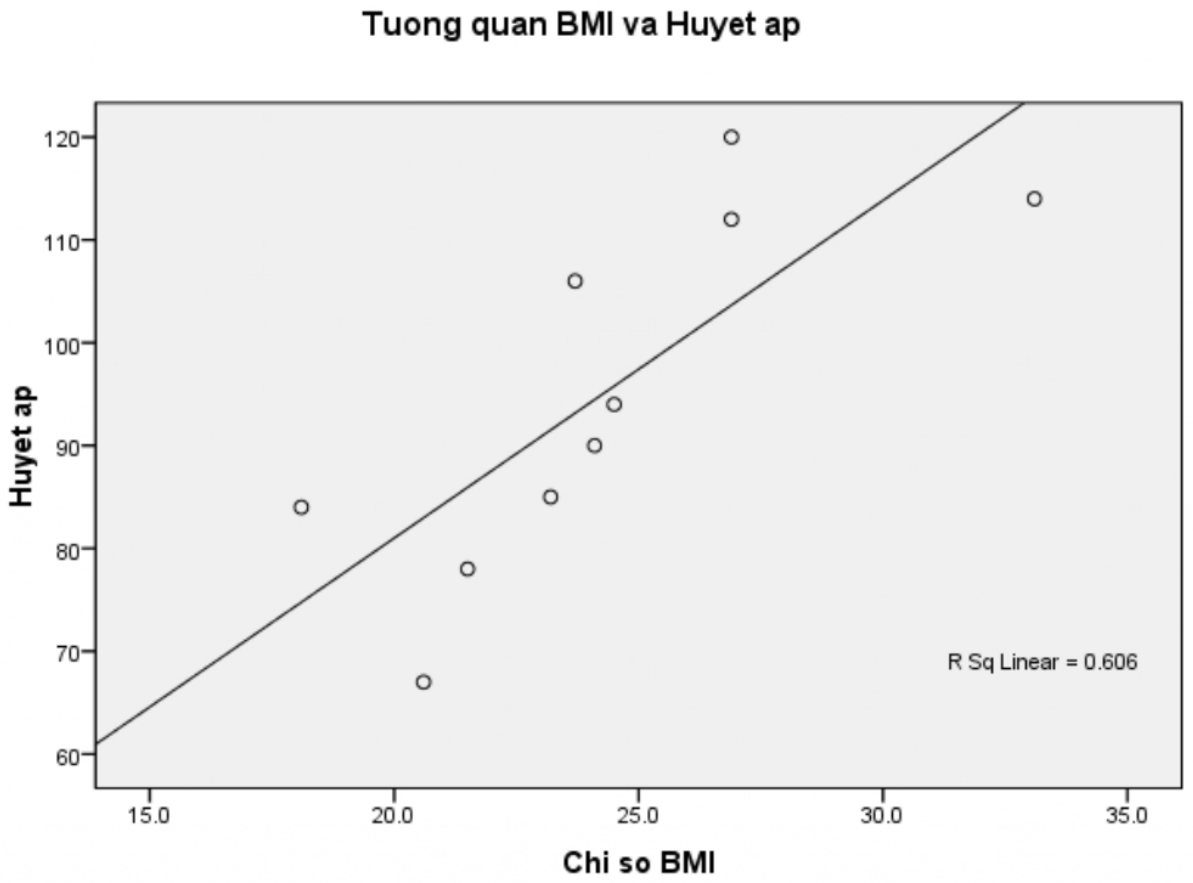
4.8. Trình bày kết quả bằng đồ thị

- Trong cửa sổ output, double click vào biểu đồ thu được, xuất hiện cửa sổ chứa biểu đồ mới có tên Chart Editor
- Click phải vào biểu đồ
- Chọn Add Fit Line at Total



Trần Thúy Hiền

4.8. Trình bày kết quả bằng đồ thị



Trần Thúy Hiền

Thank You !
Email: tthien@huemed-univ.edu.vn

Chương 5. SPSS trong Kiểm định GTTK

5.1. Cơ sở lý thuyết thống kê về ước lượng và kiểm định giả thuyết thống kê

5.2. Kiểm định độc lập (mối liên hệ) trong SPSS

5.3. Kiểm định phù hợp (dạng phân phối) trong SPSS

5.4. Kiểm định tỉ lệ trong SPSS

5.5. Kiểm định trung bình trong SPSS

5.5.1. Kiểm định trung bình bằng test tham số

5.5.2. Kiểm định trung bình bằng test phi tham số

Trần Thúy Hiền

Chương 5. SPSS trong Kiểm định GTTK

Mục tiêu

- ✓ Chọn lựa được kiểm định thống kê phù hợp cho các yêu cầu phân tích trong nghiên cứu.
- ✓ Hiểu được các giả định liên quan đến các kiểm định thống kê.
- ✓ Sử dụng được SPSS để thực hiện các kiểm định thống kê.
- ✓ Ứng dụng kiểm định giả thuyết thống kê để giải quyết vấn đề trong nghiên cứu y học, đưa ra khuyến nghị trong y học.

Chương 5. SPSS trong Kiểm định GTTK

5.1. Cơ sở lý thuyết thống kê về ước lượng và kiểm định giả thuyết thống kê

5.1.1. Ước lượng khoảng tin cậy

- ✓ Khoảng tin cậy (Confident interval) cho tham số $\mu, p, \sigma^2 \ldots$ của quần thể với độ tin cậy $(1-\alpha)$

  Độ tin cậy 90%, 95%, 99%...: $\alpha = 0,1;\ 0,05;\ 0,01 \ldots$

- ✓ Khoảng tin cậy cho tỉ lệ p của tổng thể (cỡ mẫu n lớn, $nf > 5$; $n(1-f) > 5$) là:

$$\left(f \pm u(\alpha/2).\frac{\sqrt{f(1-f)}}{\sqrt{n}} \right)$$

Chương 5. SPSS trong Kiểm định GTTK

- ✓ Khoảng tin cậy cho trung bình μ của tổng thể (tổng thể thỏa mãn phân phối chuẩn) là:

Nếu phương sai tổng thể đã biết:

$$\left(\overline{X} \pm u(\alpha/2).\frac{\sigma}{\sqrt{n}}\right)$$

Nếu phương sai tổng thể chưa biết:

$$\left(\overline{X} \pm t_{n-1}(\alpha/2).\frac{S}{\sqrt{n}}\right)$$

Chương 5. SPSS trong Kiểm định GTTK

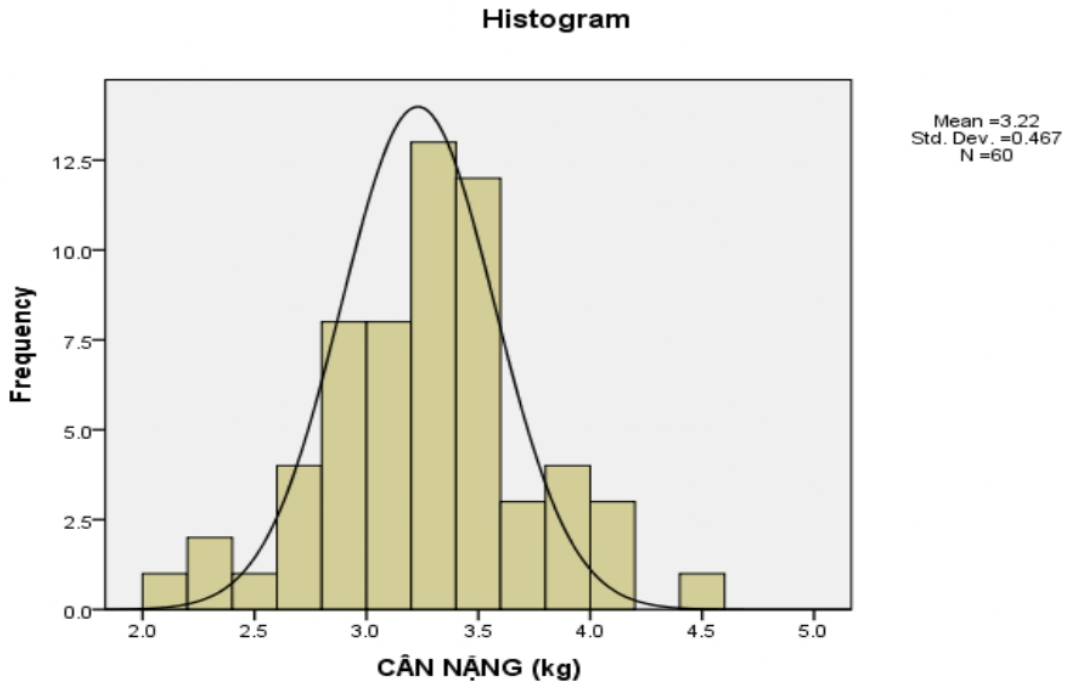
Ví dụ 5.1. Cân nặng (kg) của 60 trẻ sơ sinh được chọn ngẫu nhiên trong số trẻ sinh ra tại một bệnh viện

2,2	2,9	2,9	2,9	2,9	2,9	3,0	3,0	3,0	3,0
3,1	3,1	3,1	3,1	3,2	3,2	3,2	3,2	3,2	3,2
3,3	3,3	3,3	3,3	3,3	3,3	3,3	3,4	3,4	3,4
3,4	3,4	3,5	3,5	3,5	3,5	3,5	3,5	3,5	3,6
3,6	3,6	3,8	3,8	3,8	3,9	4,0	4,0	4,1	4,4

Dựa vào mẫu số liệu, xác định khoảng tin cậy 95% cho cân nặng trung bình trẻ sơ sinh của bệnh viện.

Chương 5. SPSS trong Kiểm định GTTK

Phân phối chuẩn:



Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
CÂN NẶNG (kg)	.089	60	.200*	.983	60	.588

a.Lilliefors Significance Correction

*.This is a lower bound of the true significance.

Chương 5. SPSS trong Kiểm định GTTK

Descriptives

			Statistic	Std. Error
CÂN NẶNG (kg)	Mean		3.217	.0603
	95% Confidence Interval for Mean	Lower Bound	3.096	
		Upper Bound	3.337	
	5% Trimmed Mean		3.222	
	Median		3.250	
	Variance		.218	
	Std. Deviation		.4673	
	Minimum		2.1	
	Maximum		4.4	
	Range		2.3	
	Interquartile Range		.6	
	Skewness		-.140	.309
	Kurtosis		.331	.608

Chương 5. SPSS trong Kiểm định GTTK

Khoảng tin cậy 95% cho cân nặng trung bình trẻ sơ sinh của bệnh viện là: (3,096; 3,337)

$$\left(\overline{X} \pm t_{59}(0,025).\frac{S}{\sqrt{n}}\right)$$

$$\left(\overline{X} \pm 1,96.Se\right)$$

Cỡ mẫu lớn: $t_{59}(0,025) \approx \mathrm{u}(0,025) = 1,96$

Chương 5. SPSS trong Kiểm định GTTK

5.1.2. Kiểm định giả thuyết thống kê

(1) Các khái niệm

- Giả thuyết thống kê là một sự xác nhận hay phỏng đoán liên quan đến một hay nhiều tổng thể.
- Giả thuyết thống kê là một mệnh đề liên quan

  - dạng phân phối xác suất $F(x,\theta)$ của biến ngẫu nhiên

  - tham số đặc trưng θ của biến ngẫu nhiên

  - tính độc lập của các biến ngẫu nhiên

Chương 5. SPSS trong Kiểm định GTTK

- Kiểm định giả thuyết là thủ tục tiến hành dựa vào thông tin của mẫu có thể đưa ra những bằng chứng để bác bỏ hay chấp nhận một giả thuyết thống kê.
  - Giả thuyết vô hiệu (null hypothesis): H_0
  - Giả thuyết thay thế (alternative hypothesis): H_1

Chương 5. SPSS trong Kiểm định GTTK

(2) Các bài toán Kiểm định – Thủ tục Kiểm định

1	2	3
Kiểm định tham số: trung bình, tỷ lệ, phương sai	Kiểm định tính độc lập	Kiểm định sự phù hợp (hay dạng phân phối)
Phép kiểm Z-test, t-test, F-test	Phép kiểm Chi square-test độc lập	Phép kiểm Chi square-test phù hợp

Mô hình Kiểm định giả thuyết của Neyman & Pearson

1) Lập giả thuyết H_0, H_1.

2) Từ mẫu ngẫu nhiên $W_X = (X_1, X_2, \ldots, X_n)$, chọn Thống kê kiểm định (Hàm kiểm định) *K* và xác định phân phối xác suất của *K* (Chọn Test thống kê)

3) Với mức ý nghĩa α, xác định Miền bác bỏ W_α

(xác định các giá trị tới hạn: $u_{\frac{\alpha}{2}}$, $t_n^{(\alpha/2)}$, $\chi_n^2(\alpha), \ldots$)

4) Mẫu số liệu: Tính giá trị quan sát K_0 (giá trị hàm *K*)

- Nếu $K_0 \in W_\alpha$ thì bác bỏ H_0, chấp nhận H_1.

- Nếu $K_0 \notin W_\alpha$ thì chưa có cơ sở bác bỏ H_0.

5) Kết luận: Ra quyết định

Mô hình Kiểm định giả thuyết trong NCYH hiện nay

Mô hình kết hợp giữa Kiểm định giả thuyết của Neyman & Pearson với Kiểm định ý nghĩa của Fisher

1) Lập giả thuyết H_0, H_1.

2) Tính cỡ mẫu theo α, β . Thu thập dữ liệu liên quan đến giả thuyết. Chọn Thống kê kiểm định *K* phù hợp (Chọn Test thống kê và kiểm tra điều kiện các giả định).

3) Mẫu số liệu: Tính giá trị quan sát K_0 và *p*-value

- Nếu $p < 0,05$ thì bác bỏ H_0, chấp nhận H_1.
- Nếu $p >= 0,05$ thì chưa có cơ sở bác bỏ H_0.

4) Kết luận: Ra quyết định

Chương 5. SPSS trong Kiểm định GTTK

- Sai lầm loại I, loại II - Mức ý nghĩa, lực kiểm định

Thực tế / Kiểm định	**H_0 sai**	**H_0 đúng**
Bác bỏ H_0	**Quyết định đúng** (*power*: lực kiểm định) $1-\beta$	**Sai lầm loại I** (*Significant level:* mức ý nghĩa) α
Chấp nhận H_0	**Sai lầm loại II** β	**Quyết định đúng** $1-\alpha$

Mô hình Kiểm định giả thuyết trong NCKH hiện nay

Kiểm định giả thuyết dùng SPSS

B1) Lập giả thuyết H_0, H_1.

B2) Chọn Test thống kê:

Thực hiện các lệnh trên SPSS:

- Chạy lệnh kiểm tra điều kiện (nếu cần)
- Chạy lệnh Test thống kê

B3) Diễn giải các Bảng phân tích test:

- Nếu Sig. < 0,05 thì bác bỏ H_0, chấp nhận H_1.
- Nếu Sig. >= 0,05 thì chưa có cơ sở bác bỏ H_0.

Kết luận: …… (đáng tin cậy không???)

Mô hình Kiểm định giả thuyết trong NCYH hiện nay

B1) Lập giả thuyết H_0, H_1.

Lập H_0

- Như nhau, giống nhau, bằng nhau, không có sự khác biệt
- Độc lập với nhau (không liên quan)
- Phù hợp với nhau

Mô hình Kiểm định giả thuyết trong NCYH hiện nay

B3) Diễn giải các Bảng phân tích test:

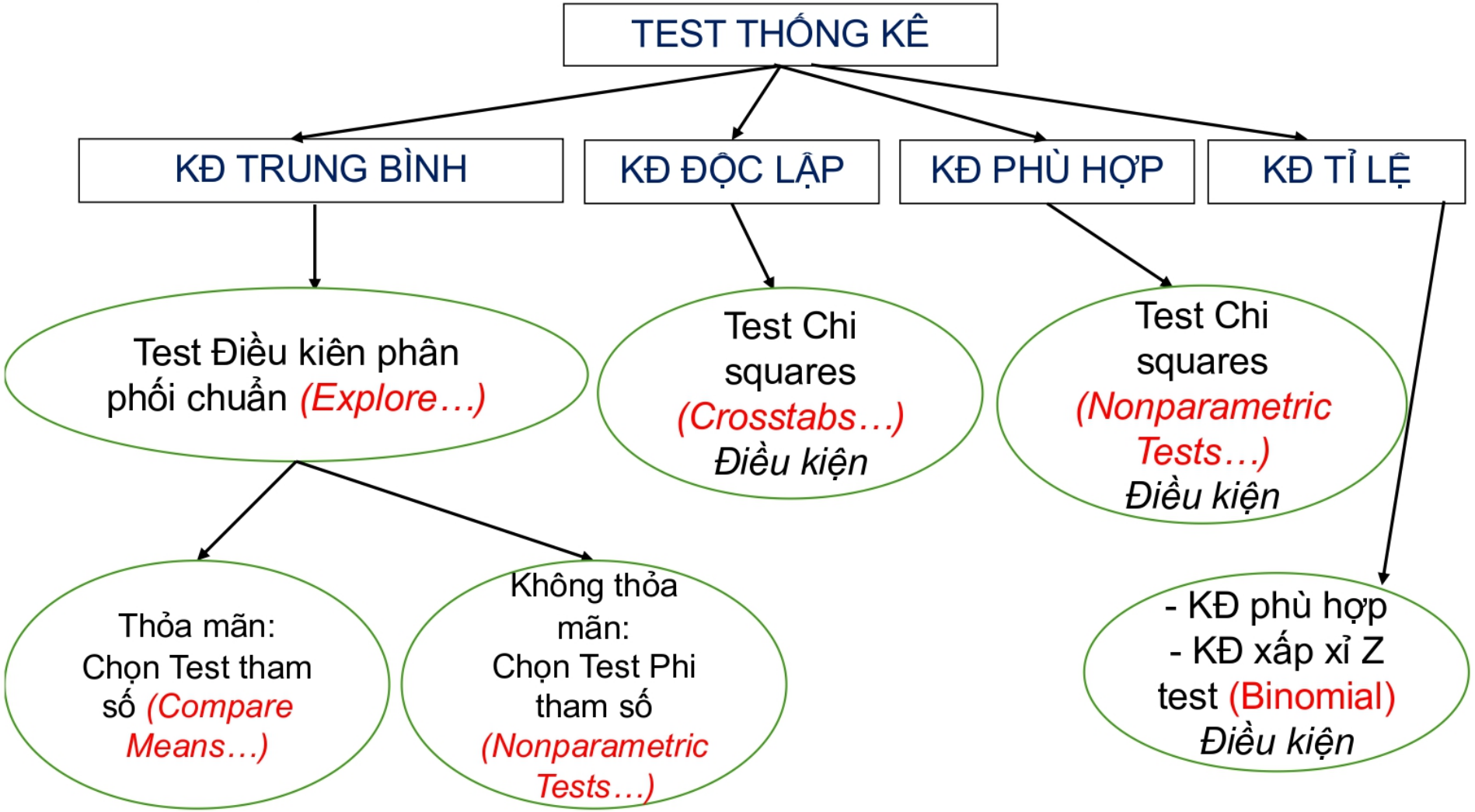
- Nếu $p < 0,05$ thì bác bỏ H_0, chấp nhận H_1.
- Nếu $p >= 0,05$ thì chưa có cơ sở bác bỏ H_0.

Kết luận: …… (đáng tin cậy không???)

SPSS: Sig.

Mô hình Kiểm định giả thuyết trong NCYH hiện nay

B2) Chọn Test thống kê: Thực hiện các lệnh trên SPSS



Chương 5. SPSS trong Kiểm định GTTK

- Lập giả thuyết thống kê H_0, H_1

Bài toán: So sánh θ và θ_0	Bài toán: So sánh θ_1 và θ_2
Lập $H_0 : \theta = \theta_0$ $\leftrightarrow \begin{cases} H_1 : \theta \neq \theta_0 & \text{(2 phía)} \\ H_1 : \theta > \theta_0 & \text{(phía phải)} \\ H_1 : \theta < \theta_0 & \text{(phía trái)} \end{cases}$	Lập $H_0 : \theta_1 = \theta_2$ $\leftrightarrow \begin{cases} H_1 : \theta_1 \neq \theta_2 & \text{(2 phía)} \\ H_1 : \theta_1 > \theta_2 & \text{(phía phải)} \\ H_1 : \theta_1 < \theta_2 & \text{(phía trái)} \end{cases}$

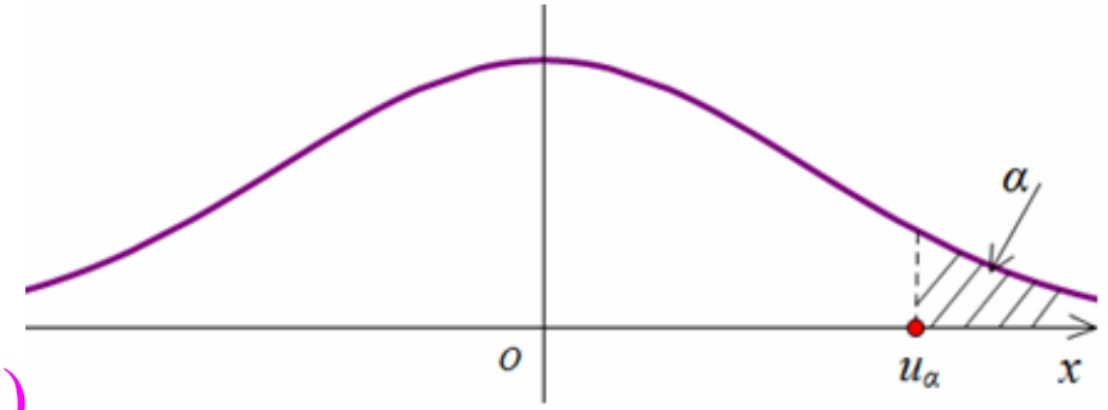
Chương 5. SPSS trong Kiểm định GTTK

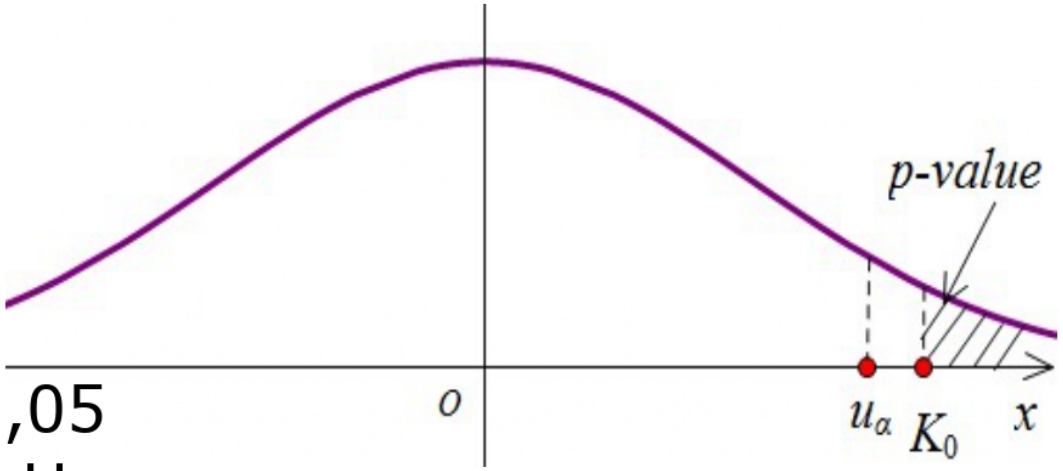
- Xác suất p – value: mức độ khẳng định việc bác bỏ H_0.

Giả sử H_0 đúng,

$$\alpha = P(K \in W_\alpha \,|\, H_0)$$

$$p-value = \begin{cases} P(K > |K_0|) \\ P(K > K_0) \\ P(K < K_0) \end{cases}$$





Thông thường: p < 0,05
thì nghiêng về bác bỏ H_0

Chương 5. SPSS trong Kiểm định GTTK

- Xác suất p – value: mức độ khẳng định việc bác bỏ H_0.

  p <<: kết quả kiểm định càng có ý nghĩa.

  p < 0,05: bằng chứng hữu ích để bác bỏ H_0

  p < 0,01: bằng chứng mạnh mẽ để bác bỏ H_0

  p >= 0,05: không có bằng chứng mạnh mẽ nào để bác bỏ H_0, sự khác biệt không có ý nghĩa thống kê.

  SPSS: Sig. (*Significant*)

Chương 5. SPSS trong Kiểm định GTTK

Chú ý, thực hiện Kiểm định GTTK trên SPSS

1) Quy tắc Lập H_0

2) Chọn Test trong SPSS:

Thao tác trên SPSS…

Ok

3) Diễn dịch kết quả trên Output: Sig. = ?

Sig. < 0,05: Có cơ sở bác bỏ H_0

Sig. >= 0,05: Chưa có cơ sở bác bỏ H_0

Chương 5. SPSS trong Kiểm định GTTK

Ví dụ 5.2. Theo một báo cáo Y tế cho biết đối với người Việt Nam lượng huyết sắc tố trung bình là 138,3 (g/l). Điều tra trên một mẫu 80 công nhân một nhà máy có tiếp xúc với hóa chất, tính toán số liệu ta có:

$\overline{X}$ = 120 (g/l) và S = 15 (g/l).

Kiểm định xem: lượng huyết sắc tố trung bình của công nhân nhà máy có tiếp xúc hóa chất có thấp hơn mức trung bình chung không? Mức ý nghĩa 0,05.

$$H_0 : \mu = 138,3$$

Chương 5. SPSS trong Kiểm định GTTK

Ví dụ 5.3. Nghiên cứu về bệnh bướu cổ ở các bé trai và các bé gái độ tuổi 8 – 12 ở một địa phương cho thấy: trong 60 bé trai có 18 bé bị bướu cổ, trong 70 bé gái có 25 bé bị bướu cổ.

Kiểm định xem: thực sự có sự khác biệt về tỷ lệ mắc bệnh bướu cổ ở 2 giới hay không? Mức ý nghĩa 0,05.

$$H_0 : p_1 = p_2$$
$$H_1 : p_1 \neq p_2$$

Chương 5. SPSS trong Kiểm định GTTK

Ví dụ 5.4. Phân bố các nhóm máu O, A, B, AB trong một cộng đồng dân cư tương ứng là 40%, 43%, 12%, 5%. Điều tra trên một mẫu 100 người dân được chọn trong vùng T thuộc cộng đồng dân cư nói trên, kết quả

Nhóm máu	O	A	B	AB
Số người	35	35	20	10

Kiểm định xem: Có sự phù hợp giữa sự phân bố các nhóm máu của vùng T với sự phân bố các nhóm máu trong cộng đồng dân cư đó hay không?

H_0: Có sự phù hợp

Chương 5. SPSS trong Kiểm định GTTK

Ví dụ 5.5. Gọi X (mmHg) là số đo huyết áp và Y (kg) là cân nặng của trẻ 14 tuổi là đối tượng nghiên cứu.

Điều tra trên một mẫu 200 trẻ, số liệu thu được

X Y	≤ 99	99 - 110	110 - 120	> 120
< 40 (kg)	10	20	11	5
≥ 40 (kg)	6	48	50	50

Kiểm định xem: Có mối liên quan giữa huyết áp và cân nặng hay không?

H_0: Không có mối liên quan

Chương 5. SPSS trong Kiểm định GTTK

(3) Ý nghĩa của OR (odds ratio) và RR (relative risk):

OR và RR là 2 chỉ số phản ánh mối liên hệ giữa một yếu tố nguy cơ và bệnh tật.

OR thường dùng trong nghiên cứu mô tả cắt ngang và nghiên cứu bệnh chứng.

RR thường dùng trong nghiên cứu thuần tập.

* Số chênh của bệnh B: khả năng mắc bệnh so với khả năng không mắc bệnh

$$odds\ (B) = \frac{P(B)}{1-P(B)} = \frac{p}{1-p}$$

Chương 5. SPSS trong Kiểm định GTTK

* Tỉ số chênh OR:

$$\text{OR} = \frac{odds_1}{odds_2} = \frac{p_1}{1-p_1} : \frac{p_2}{1-p_2}$$

Số chênh của bệnh B trong nhóm phơi nhiễm yếu tố nguy cơ: $odds_1 = \frac{p_1}{1-p_1}$

Số chênh của bệnh B trong nhóm không phơi nhiễm yếu tố nguy cơ: $odds_2 = \frac{p_2}{1-p_2}$

* Nguy cơ tương đối RR:

$$\text{RR} = \frac{p_1}{p_2}$$

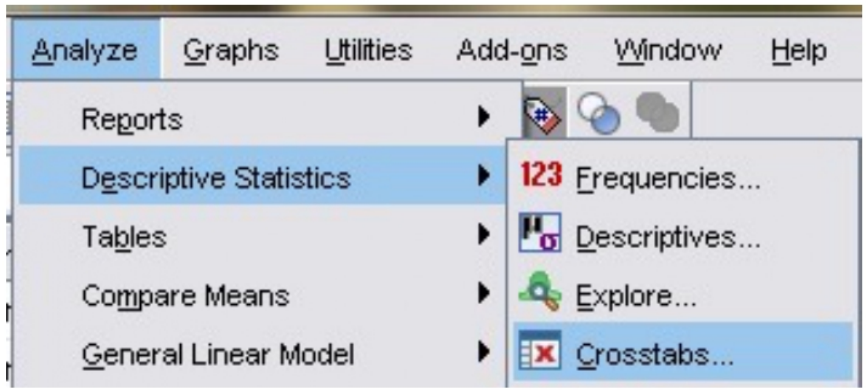
Chương 5. SPSS trong Kiểm định GTTK

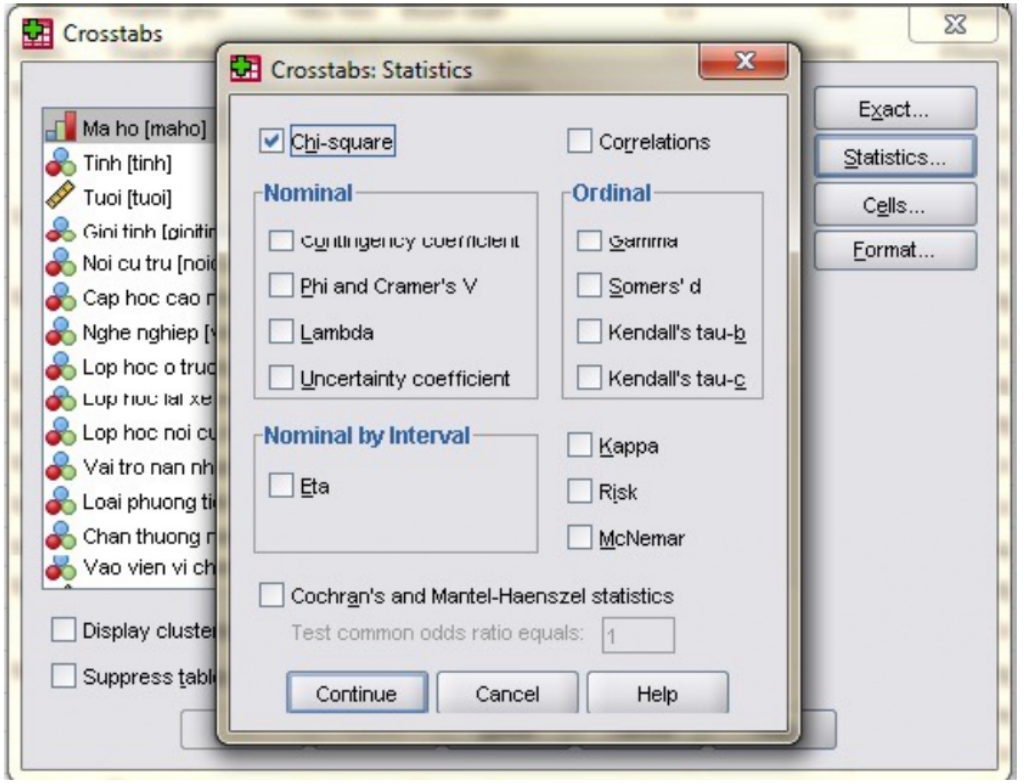
(4) Test thống kê trong nghiên cứu:

- Phép kiểm *F-test*: Thống kê kiểm định có phân phối Fisher Snedecor $F_{m,n}$
- Phép kiểm *t-test*: Thống kê kiểm định có phân phối Student t_n
- Phép kiểm *Z-test*: Thống kê kiểm định có phân phối chuẩn hóa *N(0;1)*
- Phép kiểm *Chi Square*: Thống kê kiểm định có phân phối khi bình phương χ^2_n

Test Thống kê trong SPSS

1. *Phép kiểm Chi square độc lập*

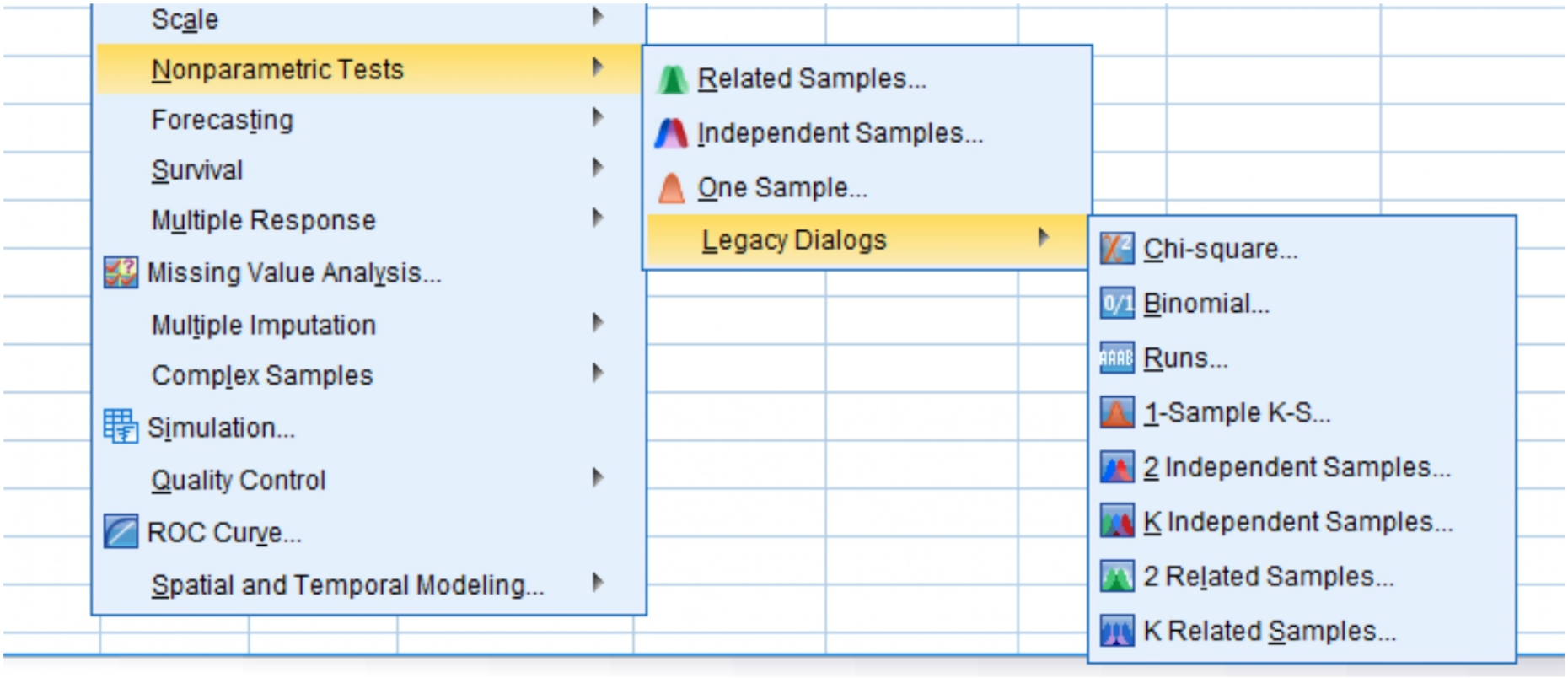




Trần Thúy Hiền

Test Thống kê trong SPSS

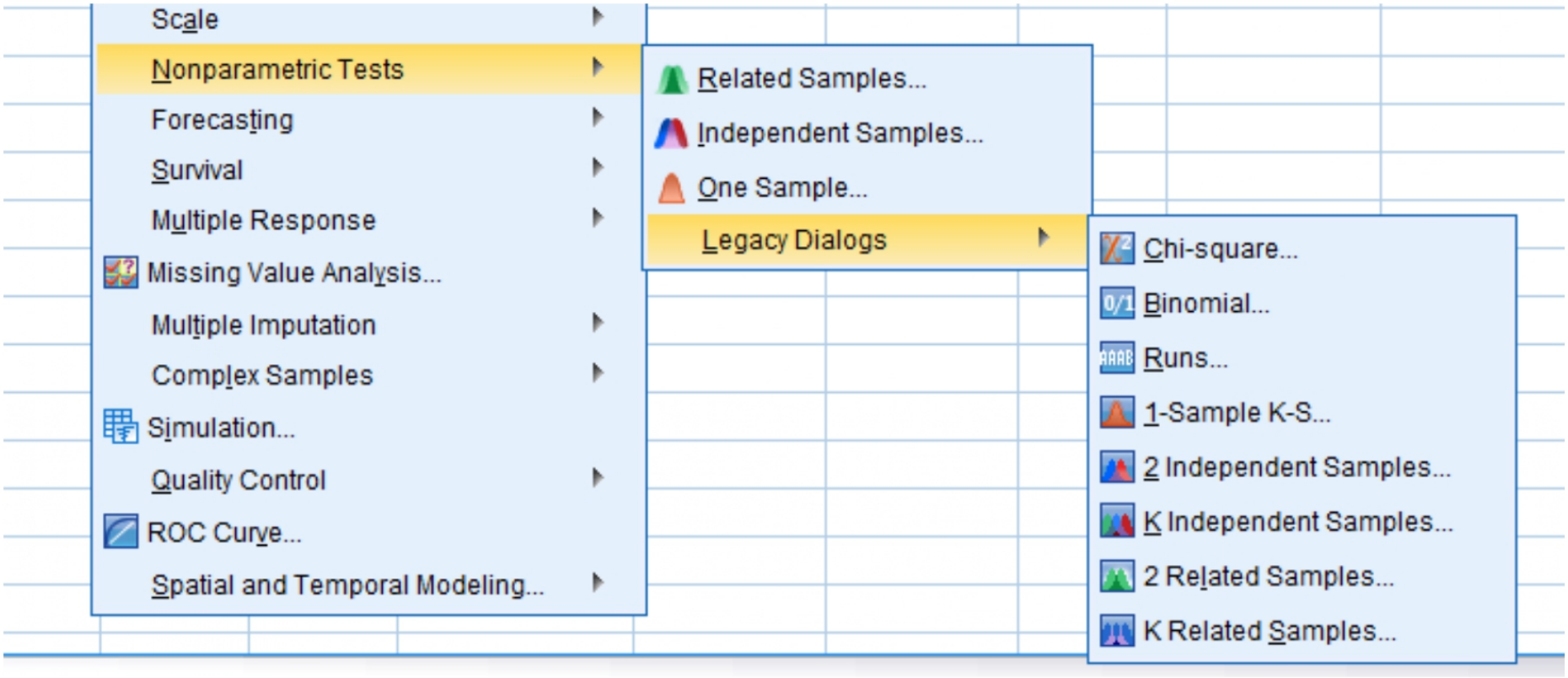
2. ***Phép kiểm Chi square phù hợp***



Trần Thúy Hiền

Test Thống kê trong SPSS

3. *Phép kiểm Binomial*



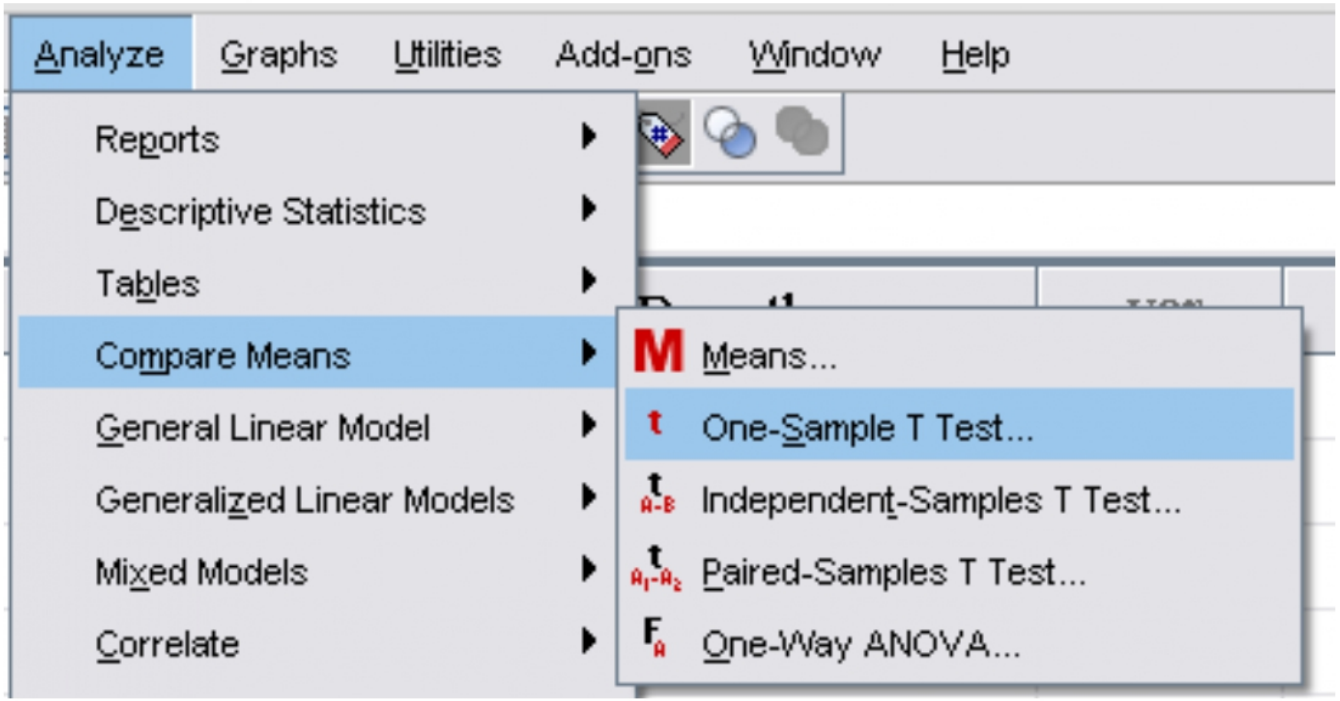
Trần Thúy Hiền

Test thống kê trong SPSS

4. ***Phép kiểm F-test***

SPSS: **Levene's Test** for Equality of Variances

5. ***Phép kiểm t-test một mẫu***



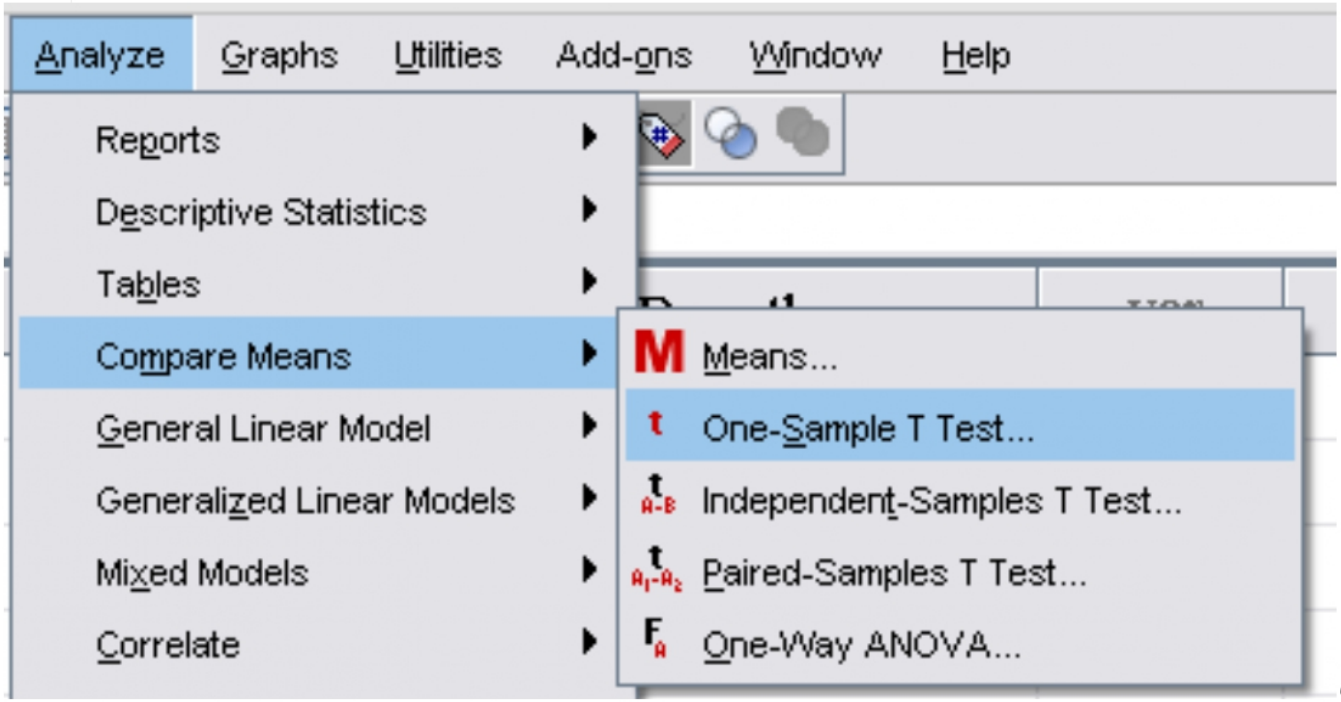
Trần Thúy Hiền

Test Thống kê trong SPSS

6. ***Phép kiểm t-test 2 mẫu độc lập***

7. ***Phép kiểm t-test mẫu ghép cặp***

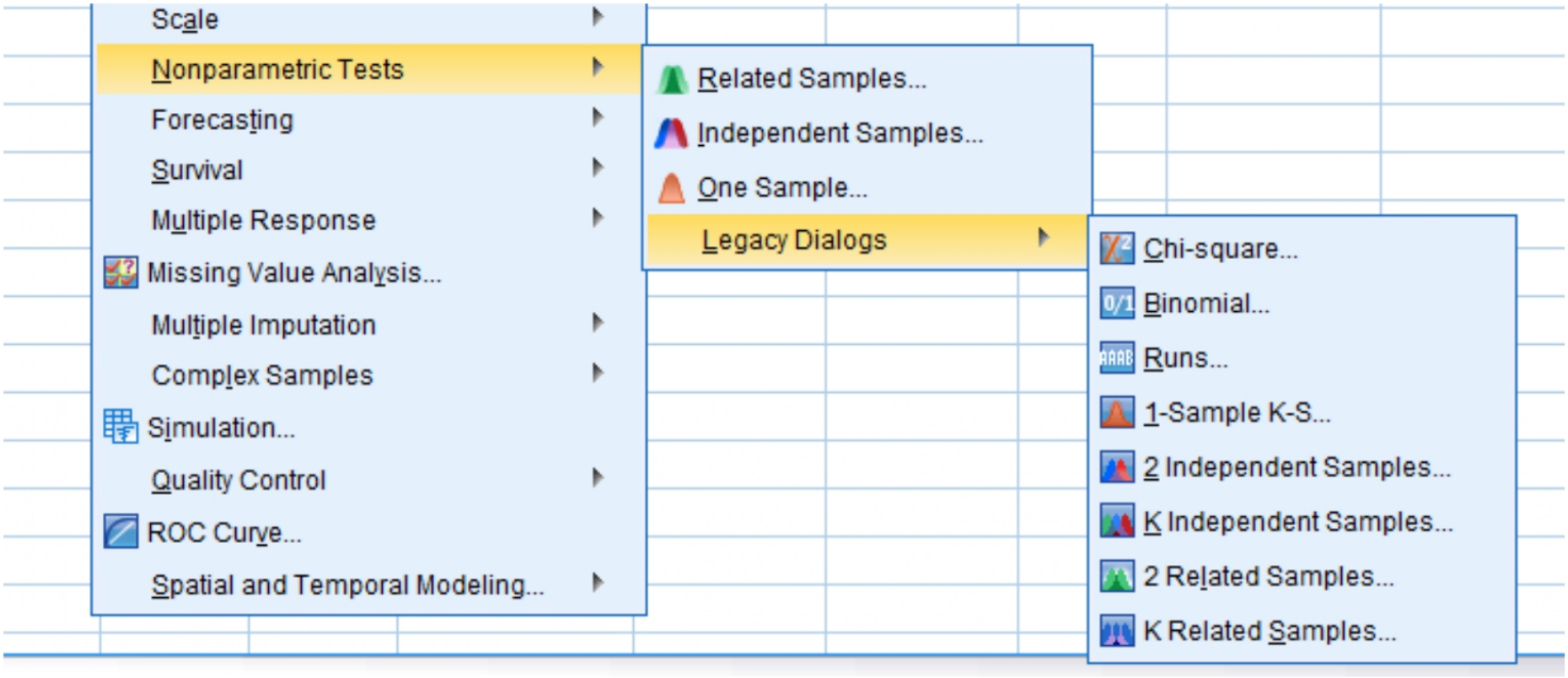
8. ***Phép kiểm ANOVA***



úy Hiền

Test Thống kê trong SPSS

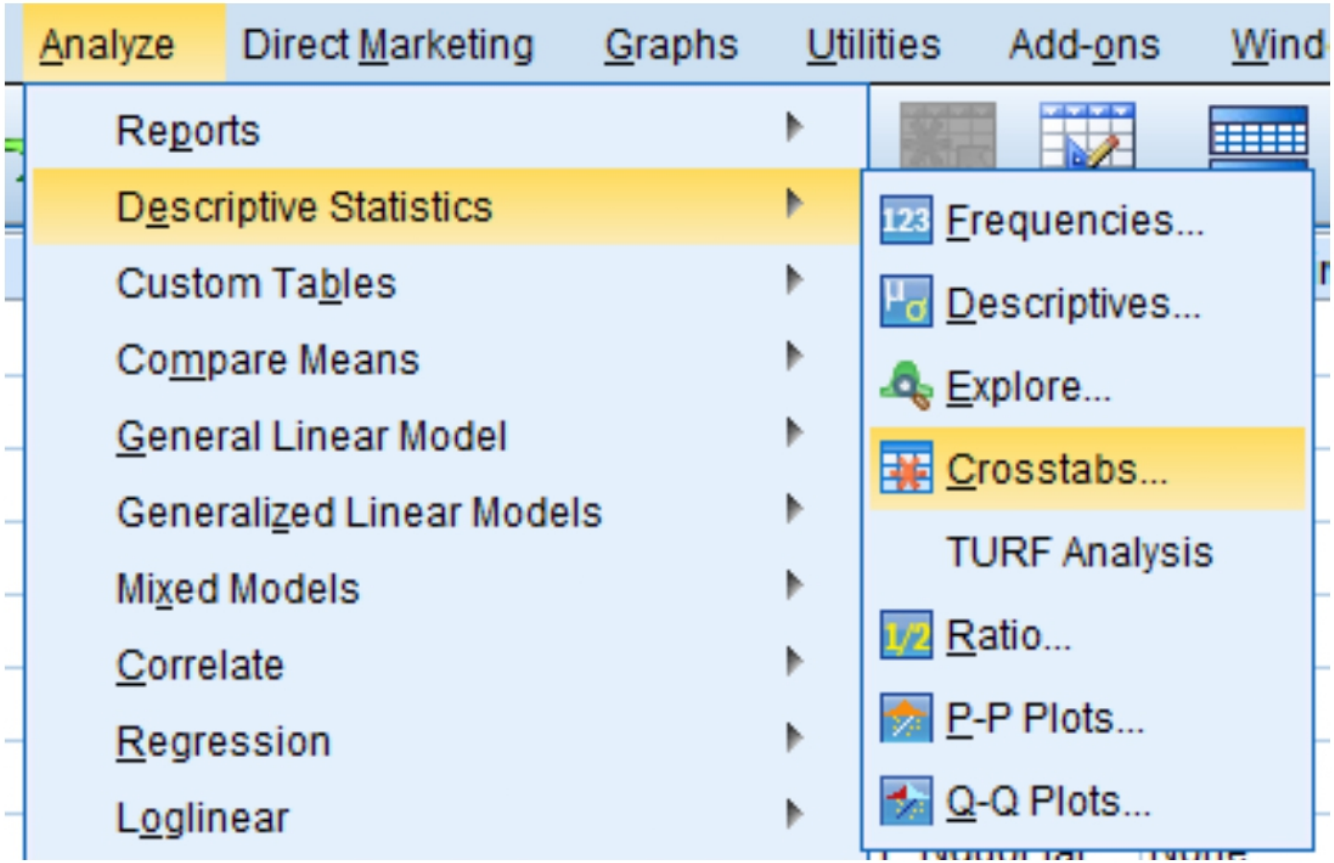
9. ***Phép kiểm phi tham số***



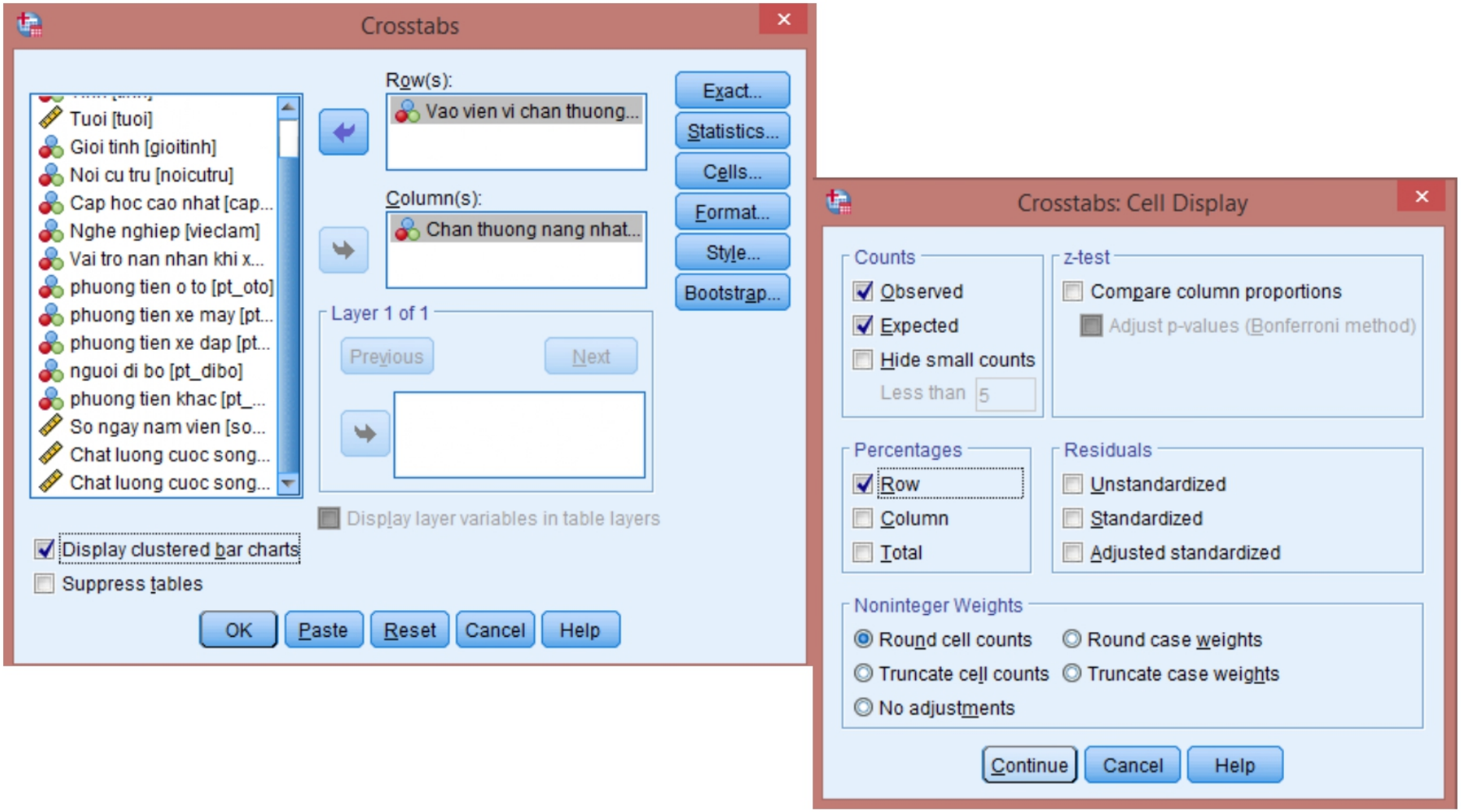
Trần Thúy Hiền

5.2. Kiểm định độc lập trong SPSS

Ví dụ 5.6. Kiểm định xem có mối liên quan giữa vị trí chấn thương nặng và việc có nằm viện hay không?



Chương 5. SPSS trong Kiểm định GTTK



Trần Thúy Hiền

Chương 5. SPSS trong Kiểm định GTTK

Vao vien vi chan thuong * Chan thuong nang nhat Crosstabulation

			Chan thuong nang nhat			Total
			Dau/Cot song	Than minh	Tay/chan	
Vao vien vi chan thuong	Khong	Count	7	8	8	23
		Expected Count	7.8	8.3	6.9	23.0
		% within Vao vien vi chan thuong	30.4%	34.8%	34.8%	100.0%
	Co	Count	10	10	7	27
		Expected Count	9.2	9.7	8.1	27.0
		% within Vao vien vi chan thuong	37.0%	37.0%	25.9%	100.0%
Total		Count	17	18	15	50
		Expected Count	17.0	18.0	15.0	50.0
		% within Vao vien vi chan thuong	34.0%	36.0%	30.0%	100.0%

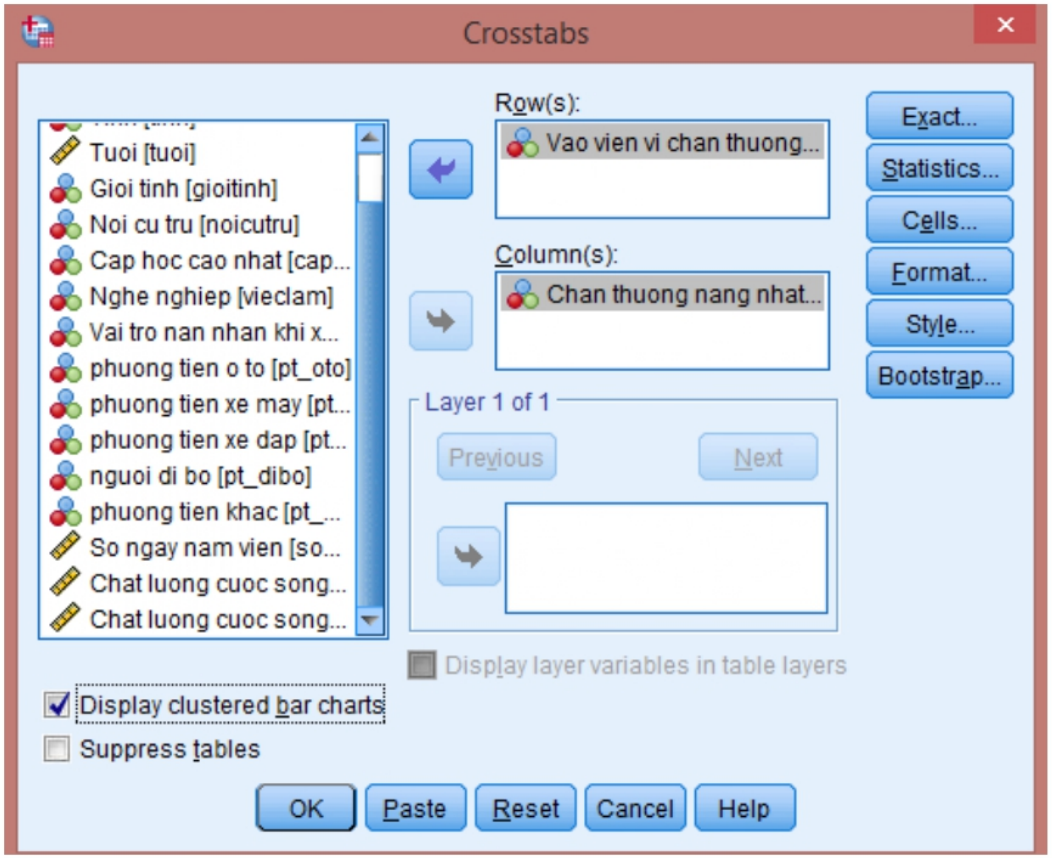
Chi-Square Tests

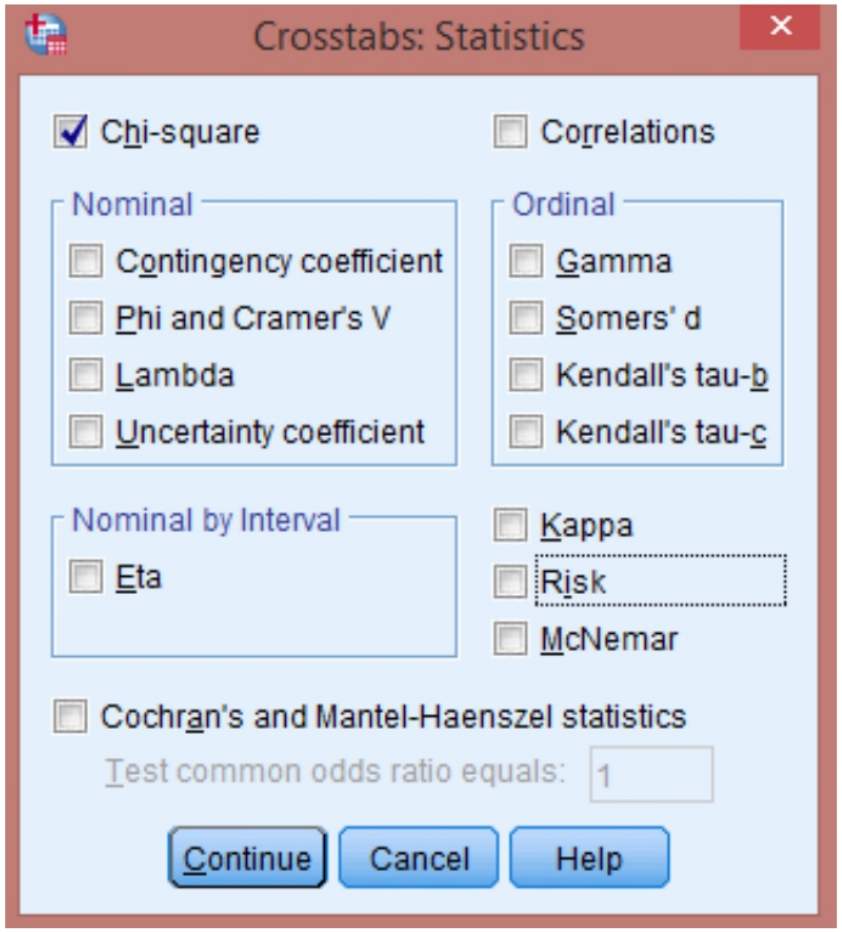
	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	.502[a]	2	.778
Likelihood Ratio	.501	2	.778
Linear-by-Linear Association	.456	1	.500
N of Valid Cases	50		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 6.90.

Trần Thúy Hiền

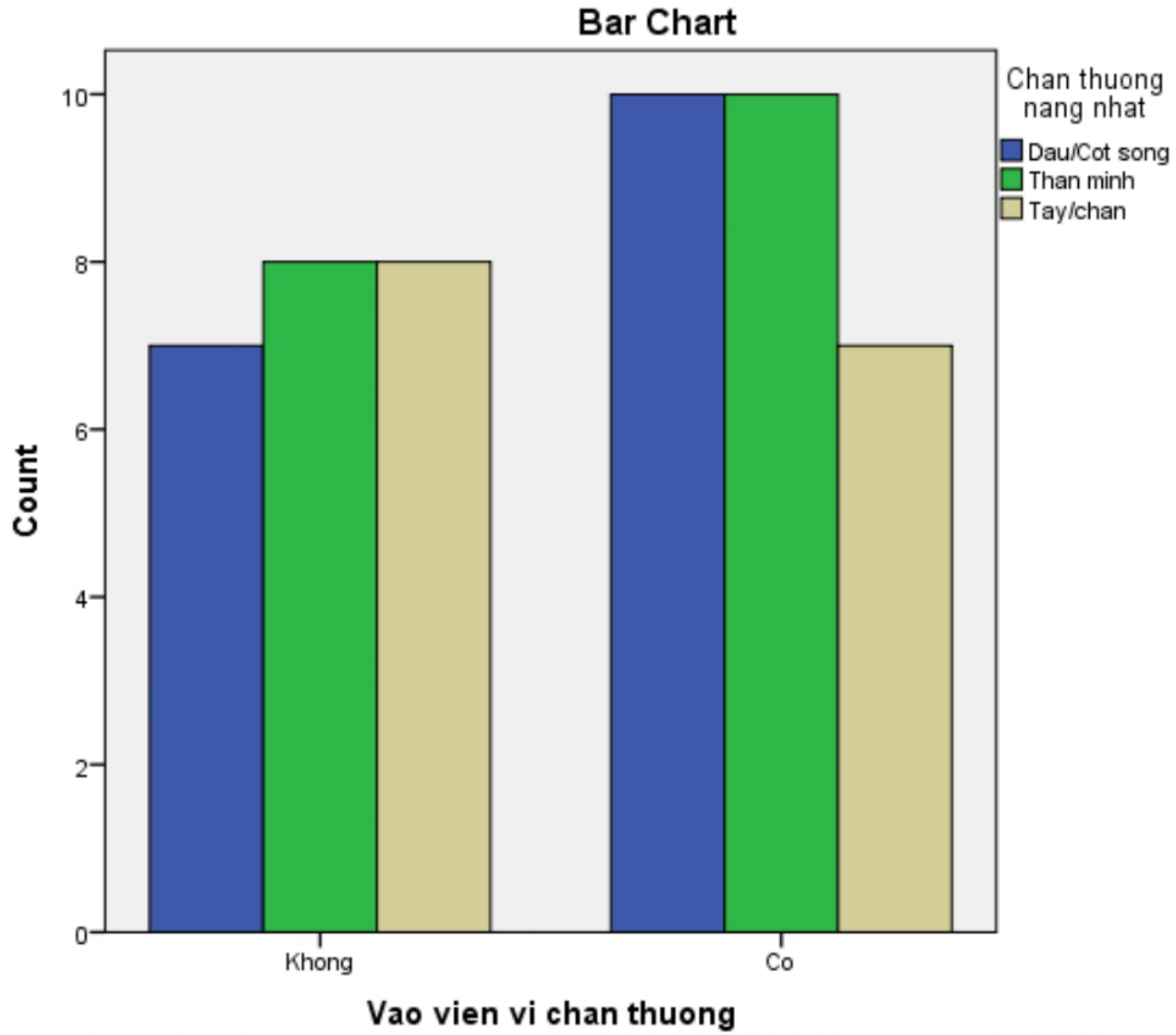
Chương 5. SPSS trong Kiểm định GTTK





Trần Thúy Hiền

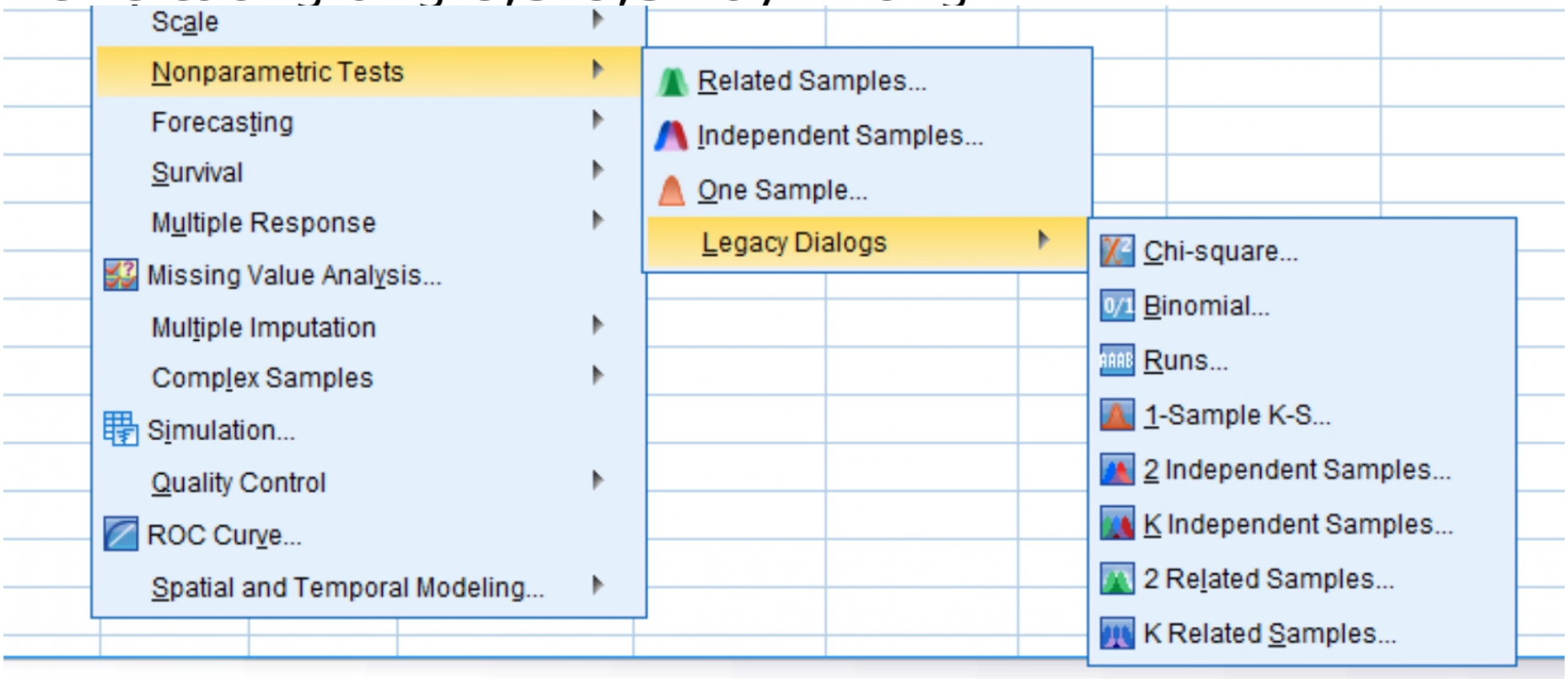
Chương 5. SPSS trong Kiểm định GTTK



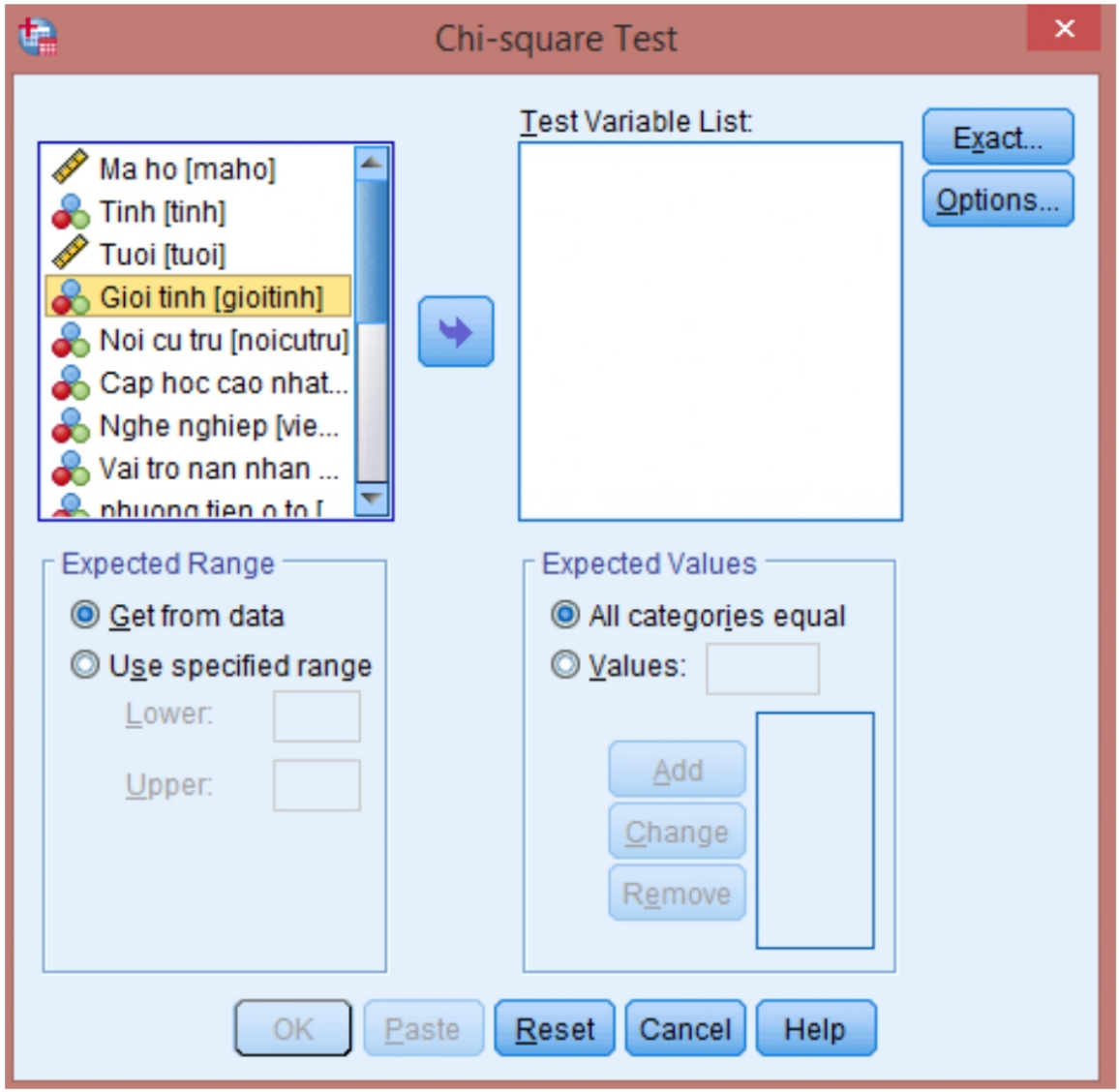
Chương 5. SPSS trong Kiểm định GTTK

5.3. Kiểm định phù hợp trong SPSS

Ví dụ 5.7. Kiểm định xem tỉ lệ nam:nữ có phù hợp với tỉ lệ tương ứng 0,5:0,5 hay không?

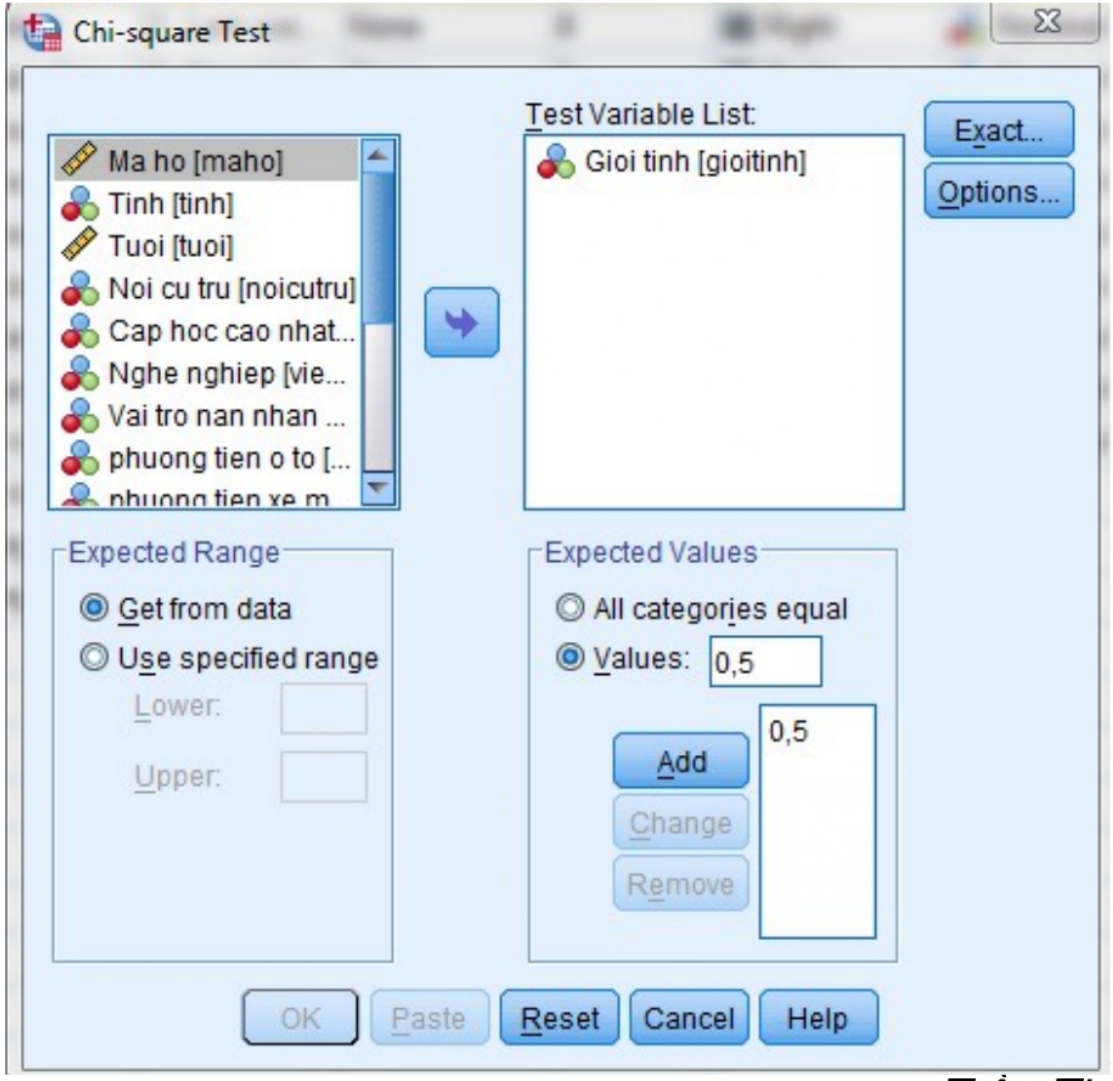


Chương 5. SPSS trong Kiểm định GTTK



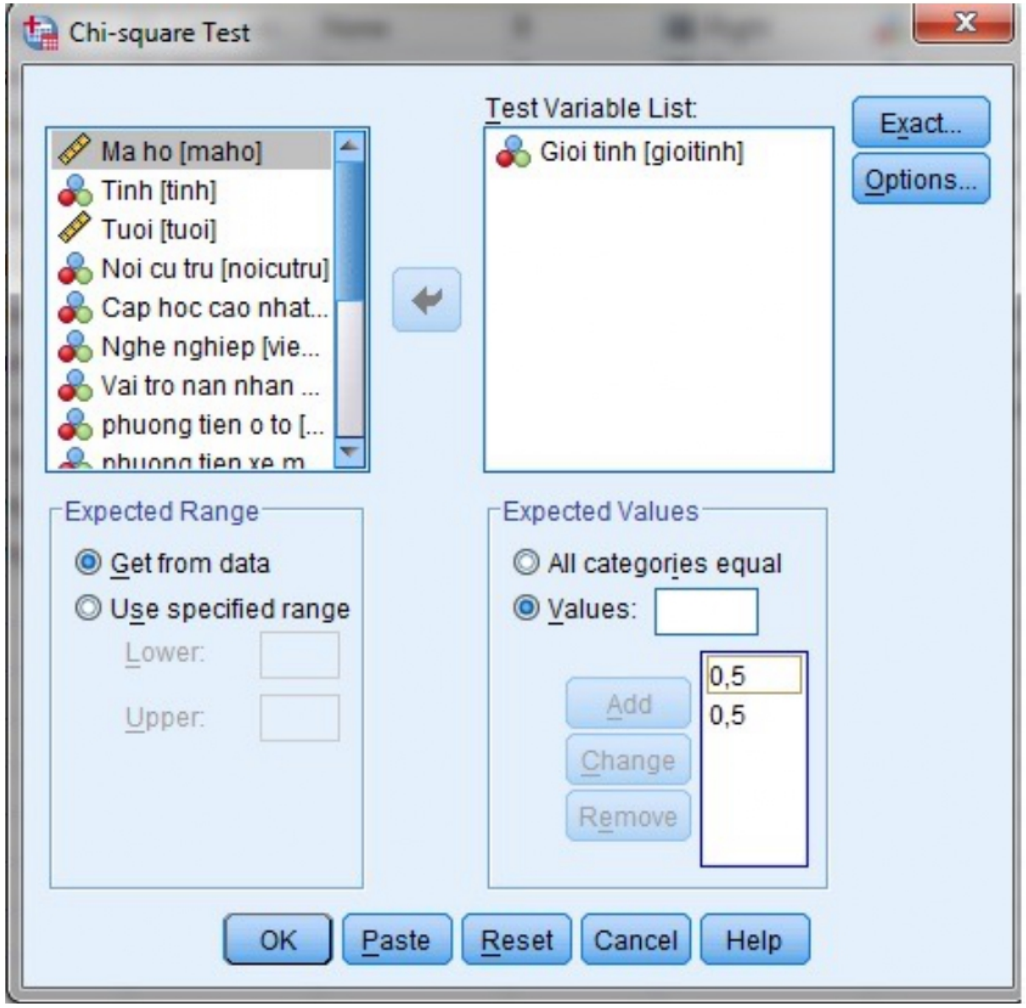
Trần Thúy Hiền

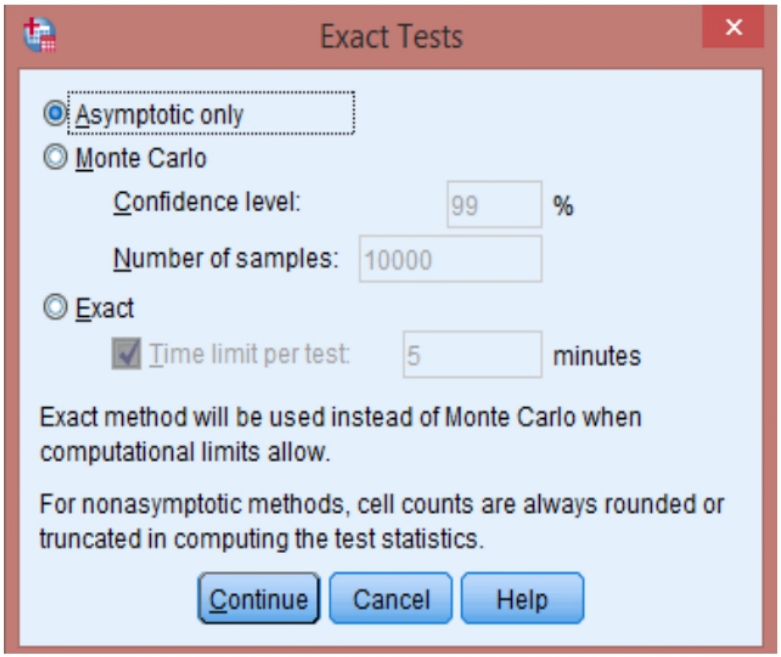
Chương 5. SPSS trong Kiểm định GTTK



Trần Thúy Hiền

Chương 5. SPSS trong Kiểm định GTTK





Trần Thúy Hiền

Chương 5. SPSS trong Kiểm định GTTK

Frequencies

Gioi tinh

	Observed N	Expected N	Residual
Nam	28	25,0	3,0
Nu	22	25,0	-3,0
Total	50		

Test Statistics

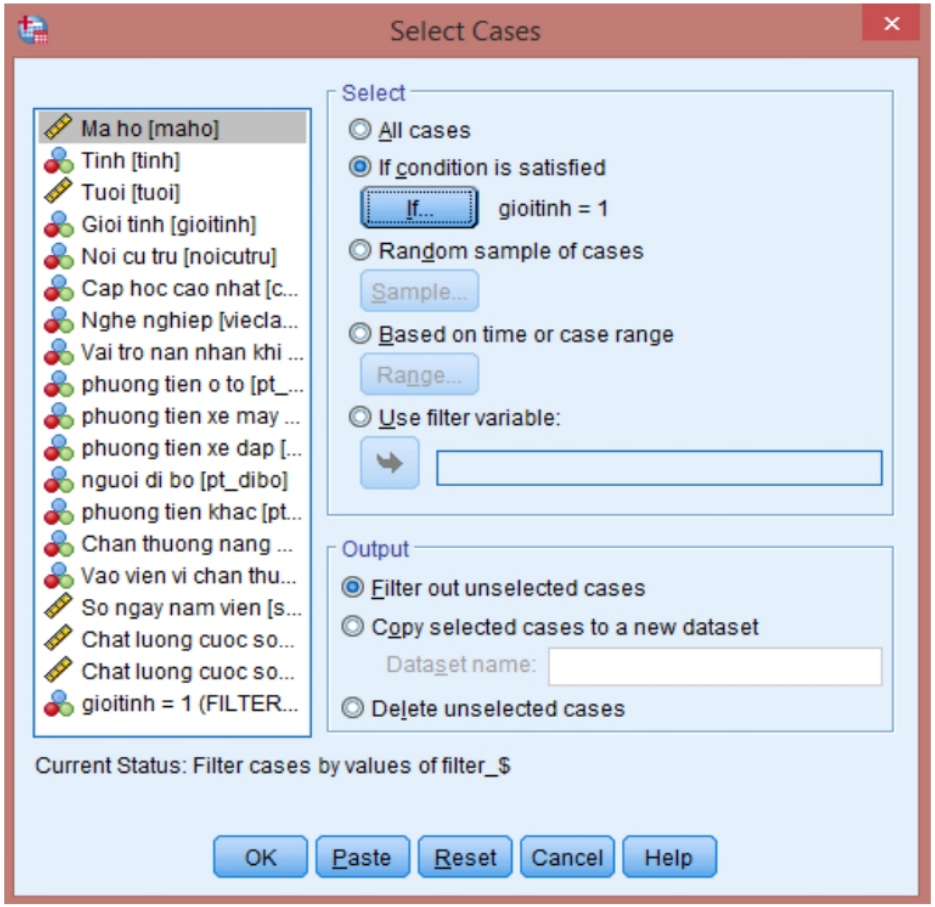
	Gioi tinh
Chi-Square	,720[a]
df	1
Asymp. Sig.	,396

a. 0 cells (0,0%) have expected frequencies less than 5. The minimum expected cell frequency is 25,0.

Chương 5. SPSS trong Kiểm định GTTK

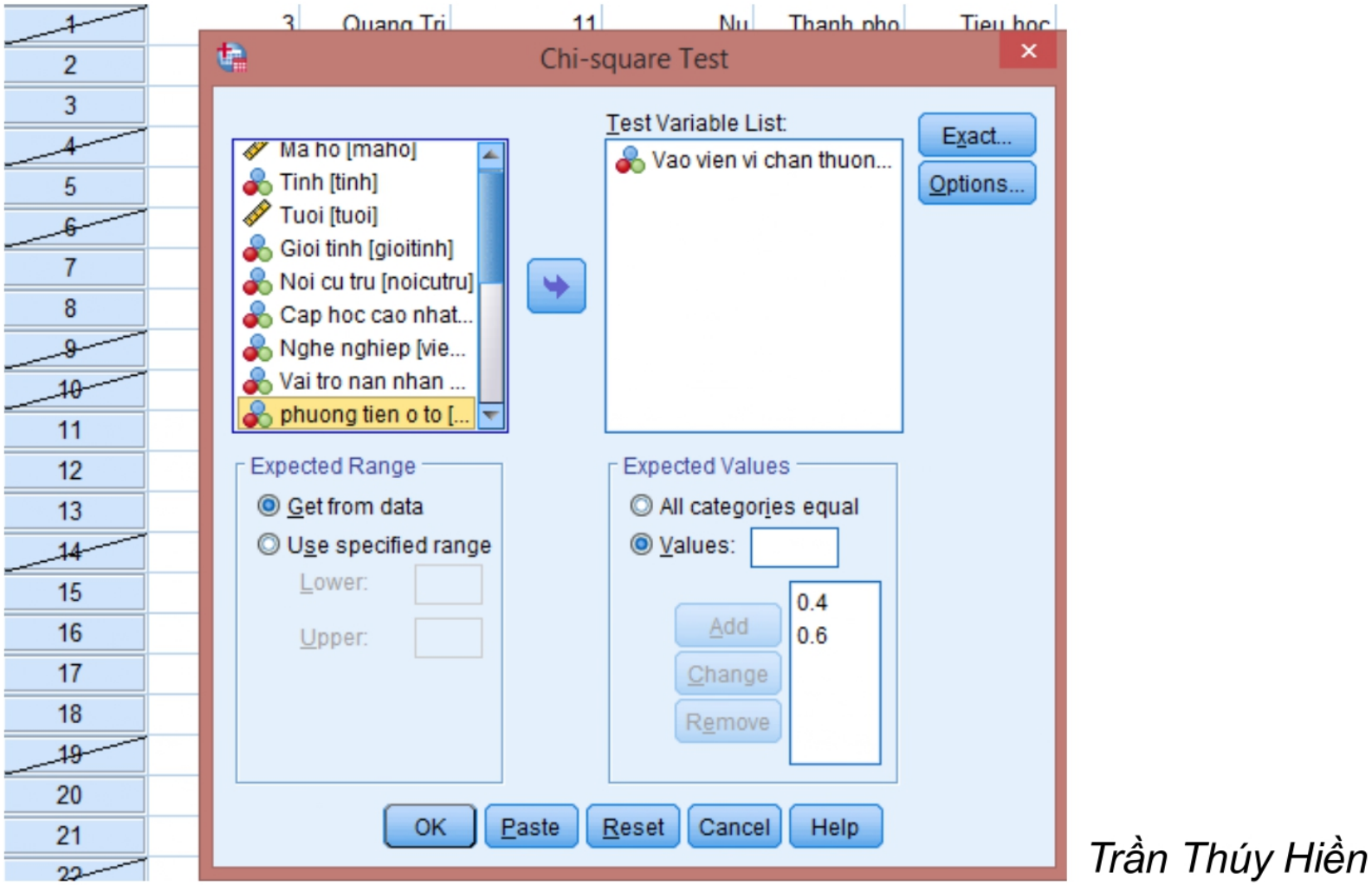
5.4. Kiểm định tỉ lệ trong SPSS

Ví dụ 5.8. Kiểm định xem tỉ lệ bệnh nhân có nằm viện vì chấn thương trong số bệnh nhân nam có thể xem là 60% hay không?



Chương 5. SPSS trong Kiểm định GTTK

Tỉ lệ bệnh nhân không nằm viện:có nằm viện là 0,4:0,6



Trần Thúy Hiền

Chương 5. SPSS trong Kiểm định GTTK

Chi-Square Test

Frequencies

Vao vien vi chan thuong

	Observed N	Expected N	Residual
Khong	13	11.2	1.8
Co	15	16.8	-1.8
Total	28		

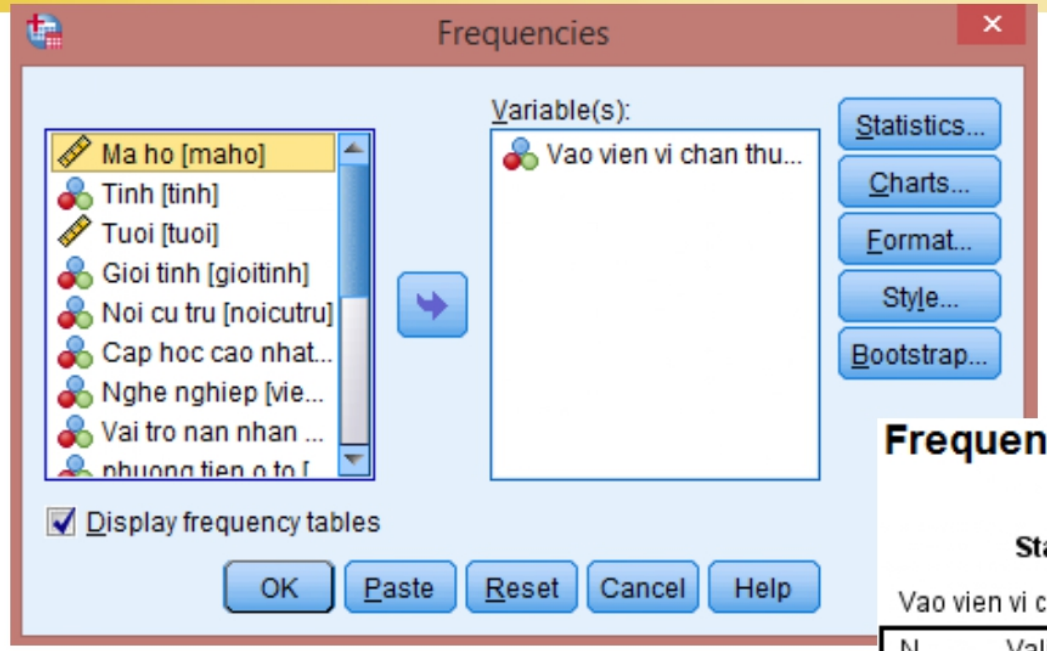
Test Statistics

	Vao vien vi chan thuong
Chi-Square	.482[a]
df	1
Asymp. Sig.	.487

a. 0 cells (0.0%) have expected frequencies less than 5. The minimum expected cell frequency is 11.2.

Trần Thúy Hiền

Chương 5. SPSS trong Kiểm định GTTK



Frequencies

Statistics

Vao vien vi chan thuong

N	Valid	28
	Missing	0

Vao vien vi chan thuong

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Khong	13	46.4	46.4	46.4
	Co	15	53.6	53.6	100.0
	Total	28	100.0	100.0	

Trần Thúy Hiền

Thank You !
Email: tthien@huemed-univ.edu.vn